Data Quality Konzept Zoo Pirmasens

Versionsnummer: 1.0  
Datum: 04.01.2024 11:49

Inhaltsverzeichnis

[1. Einleitung/Ausgangssituation](#_Toc155269173) 2

[2. Konzeptionelle Maßnahmen zur Datenqualität bei der Erstellung der Datenbank/des DataWarehouse](#_Toc155269174) 7

[3. Maßnahmen zur Sicherstellung der Datenqualität im Betrieb](#_Toc155269175) 9

[3.1 Ziele der Datenqualität](#_Toc155269176) 9

[3.2 Standards und Richtlinien](#_Toc155269177) 9

[3.3 Verantwortlichkeiten](#_Toc155269178) 10

[3.4 Prozesse und Verfahren](#_Toc155269179) 10

[3.5 Messung und Monitoring](#_Toc155269180) 10

[3.6 Problemmanagement und Eskalationsverfahren 1](#_Toc155269181)1

[3.7 Schulung und Bewusstseinsbildung 1](#_Toc155269182)2

[3.8 Kontinuierliche Verbesserung 12](#_Toc155269183)

[3.9 Compliance und rechtliche Anforderungen 13](#_Toc155269184)

[4. Zusammenfassung 1](#_Toc155269185)4

**Anhang**

Anhang 1 Maße zur Bewertung der Data Quality

Anhang 2 Allgemeine Prinzipien zur Sicherstellung der Data Quality

Anhang 3 Erstellung eines Datenqualitätsplans

# Einleitung/Ausgangssituation

Das Projekt „Digital im Zoo“ zielt darauf ab, ein integriertes IT-System für den Zoo in Pirmasens zu entwickeln. Das Projekt ist in fünf Hauptaufgaben gegliedert: Entwicklung eines Datenmodells, Umsetzung in eine relationale Datenbank, Konzeption eines Datawarehouse, Gewährleistung einer hohen Datenqualität und die Erstellung einer Präsentation.

Zur **Sicherstellung der Datenqualität** wurde hier ein Konzept entwickelt, das beschreibt, wie der Datenbestand bei Ersterfassung und fortlaufend bearbeitet wird, um dieses hohe Niveau zu erreichen und zu erhalten.

**Ausgangssituation/Aufgabe**

Der Zoo, der seit 36 Jahren besteht und kürzlich renoviert wurde, soll nun von einer papierbasierten Verwaltung auf ein digitales System umgestellt werden. Die Hauptziele sind die Verbesserung der Geschäftsprozesse und die Ermöglichung umfassender Analysen und Auswertungen.

Im Fokus stehen die Erfassung und Verwaltung von Daten über Tiere, Mitarbeiter und Futterlieferungen. Es soll eine neue Datenbankstruktur entwickelt werden, die es ermöglicht, detaillierte Informationen über Tiere, deren Unterbringung, Gesundheitszustand, Behandlungen und Futter zu speichern. Dabei werden auch die spezifischen Anforderungen an die Gehege­verwaltung und die Mitarbeiterzuordnung berücksichtigt. Für die Mitarbeiter sind detaillierte Verwaltungsprozesse vorgesehen, einschließlich der Erfassung von Vertretungen und Zuständigkeiten.

Die neue Datenbank soll die Grundlage für die operative Abwicklung des Tagesgeschäfts bilden. Alle Mitarbeiter sollen auf einen konsistenten, validen und aktuellen Datenbestand zugreifen können.

Die Übertragung der analogen Altdaten in die Datenbank ist nur insoweit vorgesehen, wie dies für die künftige Verarbeitung notwendig wird. .

Die Einführung eines neuen IT-Systems beinhaltet die Anschaffung eigener Hardware und die Einstellung von IT-Personal. Externe Dienste, wie die Lohnabrechnung, bleiben extern. Ein zukünftiger Webshop mit virtuellen Rundgängen wird ebenfalls in Betracht gezogen.

Für die Business Intelligence (BI) sind Analysen über verschiedene Dimensionen hinweg geplant, darunter Gattung, Tierart, Krankheitsverläufe und Mitarbeiter. Alle Daten sollen im Data Warehouse (DWH) historisiert werden, um zeitliche Entwicklungen nachvollziehen zu können.

Ausgangspunkt der Konzeption sind folgende **Informationen und Anforderungen:**

**Tierpfleger:**

Im Zoo sind die Tiere jeweils in passenden Gebäuden oder Gehegen untergebracht. Die Betreuung dieser Tiere obliegt den Tierpflegern, wobei ein Tierpfleger durchaus mehrere Tiere oder Tierarten betreuen kann. Es ist wichtig, eine klare Zuweisung zu haben, welcher Pfleger für welche Tiere verantwortlich ist. Für bestimmte Aufgaben, wie das Baden eines Elefanten oder das Striegeln großer Raubkatzen, werden in der Regel mehrere Pfleger benötigt. Zudem ist die Verwaltung von Urlaubs- und Krankheitsvertretungen für alle Angestellten des Zoos essenziell.

Für die **Mitarbeiter** sollen verschiedene Attribute erfasst werden, darunter:

Anrede, Titel, Geschlecht, Vorname, Nachname, Straße, Hausnummer, PLZ, Ort, Telefonnummern (Festnetz und Mobil), Faxnummer, E-Mailadresse, Beschäftigungsbeginn, Geburtsdatum, Geburtsort, Nationalität, sozialer Status, Bemerkungen, Lohnsteuerklasse und Mitarbeiternummer.

**Tiere:**

Im Zoo sind von jeder Tierart ein oder mehrere Exemplare vorhanden, wobei jedes Tier genau einer Tierart zugeordnet ist. Jedes einzelne Tierexemplar besitzt einen eigenen Namen. Wenn sich Tiere im Zoo fortpflanzen, soll eine Zuordnung zu den Elterntieren (Vater und Mutter) möglich sein.

Die vorgesehenen Attribute für die Tiere umfassen den Tiernamen, Tiernummer, Größe, Gewicht, Geburtsdatum, Sterbedatum, Herkunft, den Zeitpunkt, ab dem das Tier im Zoo ist, ob es im Zoo geboren wurde, die Namen des Vatertiers und des Muttertiers, Bemerkungen sowie das Geschlecht.

Zudem ist die Verwaltung eventueller Unverträglichkeiten zwischen den Tieren wichtig, wie zum Beispiel, dass sich Löwe 'Wolfgang' nicht mit Löwe 'Ernst-Günther' verträgt.

Nicht alle Tiere werden im Zoo Pirmasens geboren. Einige kommen aus anderen Zoos oder Wildparks. Daher soll für jedes Tier eine Art "Lebenslauf" dargestellt werden können, der die jeweilige Tierhistorie aufzeigt.

**Gehege:**

Die Unterbringung der Tiere in ihren Gehegen im Zoo wird sorgfältig geplant, wobei es für ein Tier verschiedene Arten von Unterbringungen geben kann. Dazu gehören Tag- und Nachtgehege, Sommer-Freigehege sowie Wintergebäude. Einige Tiere, wie jene im Streichelzoo, können sich frei bewegen. In einem Gebäude können sich auf verschiedenen Etagen mehrere Gehege befinden, darunter Terrarien oder Aquarien. Manche Gehege sind so gestaltet, dass sie weitläufig sind und an mehrere Gebäude anschließen.

Für die Gehege sind bestimmte Attribute vorgesehen: Die Größe in Quadratmetern oder Kubikmetern, der Name des Geheges, die Art des Geheges (zum Beispiel Terrarium, Aquarium), die Position des Geheges, ein Feld für weitere Bemerkungen und das Baujahr des Geheges.

Diese Informationen helfen, eine optimale Umgebung für die Tiere zu gewährleisten und die Übersicht über die verschiedenen Lebensräume im Zoo zu behalten.

**Tierärzte:**

Tiere im Zoo können krank werden, weshalb die Verwaltung verschiedener externer Tierärzte erforderlich ist. Jeder Tierarzt ist für bestimmte Tiere zuständig, kann jedoch auch für mehrere Tiere verantwortlich sein. Es ist wichtig, auch die Vertretungen jedes Tierarztes zu managen. Wenn ein Tier krank war, wird seine Krankengeschichte einschließlich eines Verweises auf den behandelnden Tierarzt gespeichert.

Für die Krankengeschichte der Tiere sollen verschiedene Attribute erfasst werden: Der Grund für die Behandlung, das Datum der Behandlung, ob es sich um eine meldepflichtige Krankheit handelt (ja/nein), der Befund, die Medikation, zusätzliche Bemerkungen und das Datum des Abschlusses der Behandlung.

Für die Tierärzte sind folgende Attribute vorgesehen: Titel, Anrede, Vorname, Nachname, der Name der Praxis, Adresse (Straße, Hausnummer, PLZ, Ort, Land), Telefonnummern (Festnetz und Mobil), Faxnummer, E-Mailadresse, Ansprechpartner, das Datum, seit dem sie für den Zoo tätig sind, Bemerkungen und Bankverbindung.

**Futter/Mahlzeiten/Lieferanten:**

Im Zoo wird Futter in großen Mengen von verschiedenen Lieferanten bezogen. Jeder Lieferant bietet unterschiedliche Futterarten an, wie Nassfutter, Trockenfutter, Spezialfutter, Diätfutter und Lebendfutter. Eine bestimmte Futterart kann dabei von mehreren Lieferanten bezogen werden. Die Konditionen für jedes Futter und jeden Lieferanten variieren, wobei auch Rabatte möglich sind.

Für die Lieferanten des Zoos sollen folgende Attribute erfasst werden: Titel, Anrede, Vorname, Nachname, Firmenname, Adresse (Straße, Hausnummer, PLZ, Ort, Land), Telefonnummern (Festnetz und Mobil), Faxnummer, E-Mailadresse, Ansprechpartner, das Datum, seitdem sie für den Zoo tätig sind, Bankverbindung, Umsatzsteuernummer und Bemerkungen.

Im Rahmen der Modernisierung des Zoos ist vorgesehen, mehrere dezentrale Futterlager einzurichten. Die Verwaltung dieser Lager umfasst die Überwachung der aktuellen Bestände, das Festlegen von Meldebeständen sowie die Erfassung von Datum und Uhrzeit der Bestandsaufnahmen.

Die Tiere im Zoo werden mit dem eingekauften Futter gefüttert, wobei die Fütterungen nach einem festen Zeitplan stattfinden. Sie sind für bestimmte Tage und festgelegte Uhrzeiten angesetzt. Bei den Fütterungen werden sogenannte 'Mahlzeiten' verabreicht, die aus einer Mischung verschiedener Futtersorten bestehen können.

**Tierarten/Gattungen:**

Für alle im Zoo gehaltenen Tierarten sollen umfassende Informationen verwaltet werden, da Zoobesucher oft nach detaillierten Auskünften über die Tiere suchen. Diese Informationen umfassen den natürlichen Lebensraum, Verhaltensweisen, Abstammung, verwandte Arten und weitere relevante Details.

Es ist wichtig, dass diese Daten für jede Tierart sorgfältig gepflegt und aktualisiert werden.

Zusätzlich sollen allgemeinere Informationen bei den jeweiligen Gattungen geführt werden. Diese Gattungsinformationen können Angaben enthalten, die für alle zu dieser Gattung gehörenden Tierarten relevant sind. Einer Gattung können mehrere Tierarten zugeordnet sein. Jede Tierart gehört genau zu einer Gattung.

**Rundwege:**

Für die Zoobesucher sind speziell farblich markierte Rundwege vorgesehen, die jeweils ein bestimmtes Thema oder eine bestimmte Tiergruppe hervorheben. Beispiele für solche Rundwege könnten "Weg 1: Raubtier-Fütterungstour", "Weg 2: Gefiederte Freunde" oder "Weg 3: Alles Giftige" sein.

Infoblätter hierfür soll es kostenlos am Eingang geben. Barrierefreie Rundwege sollen gekennzeichnet sein. Die Länge in Meter sowie die Gehdauer in Minuten pro Rundweg soll abgebildet werden können.

Die Datenbank soll folgende **zentrale Geschäftsvorfälle** abbilden:

* Neues Tier anlegen / in diesem Zuge ggf. auch neue Gattung / Tierart anlegen
* Unterbringung der Tiere
* Geburt von Jungtieren
* Krankheit bei Tieren / Krankheitsverlauf
* Behandlung durch Ärzte
* Ärztevertretung im Krankheitsfall
* Ärztevertretung bei Urlaub
* Neuen Mitarbeiter anlegen
* Mitarbeiter Zuständigkeit für Tierart
* Mitarbeiter konkrete Zuordnung zu Tier
* Mitarbeitervertretung im Krankheitsfall
* Mitarbeitervertretung bei Urlaub
* Fütterung der Tiere
* Bestellung Futter
* Lagerung Futter, Bestandsverwaltung
* Neuen Futterlieferant anlegen

**Analysen** der Daten für die Geschäftsleitung und das Management sollen für die folgenden Dimensionen möglich sein:

* Gattung
* Tierart
* Tier
* Krankheit / Krankheitsverlauf
* Mitarbeiter
* Gehege / Gehegebelegung
* Futterlieferant
* Futterart
* Futter
* Futterverbrauch
* Lager / Lagerbestand
* Tierarzt
* Zeit
* Rundweg

# Konzeptionelle Maßnahmen zur Datenqualität bei der Erstellung der Datenbank/des DataWarehouse

Das Data Quality Konzept für den Zoo Pirmasens umfasst mehrere Phasen und Herausforderungen, die im Zuge der Digitalisierung von der bisher analogen und oft unvollständigen sowie inkonsistenten Datenverwaltung zu bewältigen sind. Im Mittelpunkt steht die Erstellung einer Datenbank für den operativen Bereich, die sukzessive auf andere Bereiche des Zoos ausgedehnt wird. Dabei ist zu berücksichtigen, dass die Mitarbeiter über wenig oder keine IT-Erfahrung verfügen.

Um eine hohe Datenqualität zu gewährleisten, wurden bei der Modellierung und Erstellung der Datenbank die folgenden Maßnahmen vollzogen:

Bei der Datenbankmodellierung wurde ein Kompromiss zwischen einem hohen Normalisierungsgrad und einer einfachen Bedienbarkeit der Datenbank gefunden, um Redundanzen vorzubeugen.

Die erstellte Datenbank wurde mit einem halbautomatischen Verfahren mit dem ERM abgeglichen, um die Übereinstimmung der Datenbank mit dem auf Grundlage der Kundeninformationen erstellten Modell zu gewährleisten.

Zur Datenbank liegt ein „Data Dictionary“ vor, das den Datenbankanwendern die richtige Verwendung der Datenbank erlauben soll. Insbesondere beim Eintragen und Interpretieren von Datensätzen sowie bei der Verarbeitung von Daten sollten sich die Verwender mit der Bedeutung der einzelnen Datenfelder (Attribute) vertraut machen. Auch der Datenbankmanager sollte sich beispielsweise zum Erstellen von Überprüfungen und Werteeinschränkungen mit dem „Data Dictionary“ vertraut machen.

Es wurden verschiedene Referenztabellen für Wertebereiche von Attributen (z.B. Vertretungsgründe) angelegt und eingelesen, sowie für Konsistenzbedingungen (bisher nur PLZ-Ort). Bei der Frontend-Entwicklung sollen diese verwendet werden, um Überprüfungen und Einschränkungen für diese Attribute anzulegen. Beim Einlesen der Postleitzahlen wurden vierstellige Postleitzahlen zum fünfstelligen Standard hin korrigiert.

Im Entwicklungsprozess standen die Modellentwicklung und die Überprüfung, inwieweit die Kundenanforderungen umgesetzt wurden, im Vordergrund. Es mussten relevante und für die Beteiligten verständliche Entitäten, Attribute und Relationen definiert werden. Auch im weiteren ist enge Einbindung aller Beteiligten erforderlich.

Für die Ersterfassung ist zu planen, in welcher Reihenfolge Daten, wie Tiere, Futter und Lieferanten, aufgenommen werden. Die Sicherstellung einer korrekten Dateneingabe kann über weitere vordefinierte Daten mit Wertebereichen und Auswahlmöglichkeiten erfolgen, unterstützt durch das Vier-Augen-Prinzip. Regelmäßige Kontrollen der Datenqualität durch Stichproben sind im Ersterfassungsprozess vorzusehen, dies gilt auch für eine Erfassung durch Externe. Fehlende kritische Daten bei Tieren, Verträglichkeit, Gehegen, Pfleger .. können zu Gefährdungen durch die Tiere für Mitarbeiter und Besucher führen oder den Tieren schaden.

Mittels eines Data Profilings sollte nach der Ersterfassung auf fehlende Werte, doppelte Einträge, Plausibilität, Adressen, internationale Schreibweisen geprüft werden. Die Datenbereinigung kann automatisiert oder manuell erfolgen und sollte auch von externen Dienstleistern unterstützt werden. Auch hier können vordefinierte Datentypen und Wertebereiche helfen.

Eine ausreichende und verständliche Dokumentation des Modells für die Mitarbeiter ist ebenso wichtig wie die Festlegung und Schulung geeigneter Mitarbeiter und die Definition von Verantwortlichkeiten, fördert die Akzeptanz und Umsetzung der Digitalisierung (s. 3.7).

Benutzeroberflächen für die konsistente Dateneingabe sowie die Definition von Usecases und Tests sind zu erstellen, anhand der relevanten Geschäftsvorfälle.

# Maßnahmen zur Sicherstellung der Datenqualität im Betrieb

Im laufenden Betrieb werden laufend neue Daten erfasst, die Qualitätsanforderungen genügen müssen. Weiterhin „altern“ Daten, d.h. die Qualität des Datenbestandes ist aufgrund von Änderungen der realen Welt zu überprüfen.

Inkonsistenzen und Fehler im Datenbestand können dabei zu Gefahr für und durch die Tiere führen, wenn sie falsch untergebracht werden oder nicht adäquat versorgt werden. Fehlende Daten, wie Unverträglichkeiten unter Tieren, führen gegebenenfalls zu Verletzungen und Tierverlust.

Datenverlust durch technische Ausfälle, Fehlbedienung, Kriminalität ist vorzubeugen.

Alle relevanten Datenschutzgesetze und -vorschriften sind einzuhalten.

Für folgende Themen schlagen wir Maßnahmen vor:

## Ziele der Datenqualität

Bereits vorgegeben wurde eine „Over-all“- Kennzahl an Datengenauigkeit von 97%. Diese ist weiter zu spezifizieren, um der unterschiedlichen Relevanz von Datenbereichen Genüge zu tun. Für die Krankenakten der Tiere z.B. wäre eine höhere Datenqualität erwartbar, in anderen Bereichen kann temporär eine niedrigere Qualität tolerierbar sein. Aktualisierungsvorgaben und das Bereinigungsmanagement können daraufhin angepasst werden.

## Standards und Richtlinien

Neben den bereits implementierten Einschränkungen und Vorgaben für die Erfassung von Daten sind unternehmensintern Richtlinien einzurichten und weiterzuentwickeln. Um die Konsistenz von Daten zu erreichen und deren korrekte Weiterverarbeitung sicherzustellen, sind Vollständigkeit und Einheitlichkeit der Daten unabdingbar. So ist eine einheitliche Erfassung insbesondere im Bereich der Namen und Adressen auf Konventionen angewiesen, die von allen erfassenden Mitarbeitenden akzeptiert und genutzt werden.

Hierzu ist eine Vorgabe in einer schriftlichen Richtlinie notwendig, die einer laufenden Überprüfung und Erweiterung bedarf. Diese Richtlinie ist ebenfalls allen erfassenden Mitarbeitern zur Verfügung zu stellen.

Vorgaben sind auch für die zeitnahe Erfassung von Daten notwendig, da ansonsten nicht entsprechenden Entscheidungen (z.B. Futternachbestellungen) getätigt werden.

Für einige Datenbereiche wurden bereits Wertelisten erstellt, um Fehleingaben zu verhindern (z.B. Länder, Postleitzahlen). Soweit dies auch in anderen Bereichen sinnvoll erscheint (z.B. Lebensräume) können diese von den Fachbereichen entwickelt und in die Datenbank implementiert werden.

## Verantwortlichkeiten

Auf Ebene der Geschäftsleitung ist die Verantwortlichkeit für die Datenverfügbarkeit und -qualität im Ganzen einem der Leitenden zuzuordnen. Neben der grundsätzlichen Verantwortlichkeit für die IT-seitige Betreuung der Datenbank sind die Fachbereiche und damit die Endbenutzer in die Verantwortlichkeit für die enthaltenen Daten einzubeziehen.

Die Verwaltung der Datenbank beinhaltet eine nutzungsorientierte, aber die Verantwortlichkeiten widerspiegelnde Rechteverwaltung. Insbesondere für datenschutzrelevante Bereiche wie die Stammdatenverwaltung bei den Mitarbeitenden, aber auch für besonders kritische Datenbereiche (z.B. Krankenakte) ist eine restriktive und auf fachliche Kompetenz ausgerichtete Benutzerrechtsvergabe vorzusehen. Die jeweiligen Lese- und Schreibrechte sind auf die Notwendigkeit zu minimieren.

Für Kontrollen und Audits sind spezielle Nutzerkonten ohne Schreibrechte vorzusehen.

Die Verantwortung für die Aktualisierung der Dokumentation ist ebenfalls festzulegen, und sollte einer regelmäßigen Überprüfung unterzogen werden (z.B im Rahmen von Audits, s. 3.5).

Die Fachbereichsverantwortlichen sind auch für die Veranlassung und korrekte Umsetzung von Änderungen in der rechtlichen Sphäre (s. 3.9) zuständig. Eine Priorisierung verschiedener Projekte ist auf Ebene der Geschäftsleitung vorzunehmen.

## Prozesse und Verfahren

Insbesondere der Prozessablauf bei der Erfassung von Daten sowie bei der Fehlerbehandlung ist mit den Fachbereichen zu klären, und schriftlich detailliert und datenfeldbezogen festzuhalten. Eine einheitliche Vorgehensweise erleichtert dabei die Erfassung und stellt die Konsistenz von Daten sicher. Der einfache Zugang zu den grundlegenden Dokumenten wie dem „Data Dictonary“ ist für alle Endnutzer ebenfalls zu ermöglichen.

Soweit eine unterstützte Erfassung von Daten möglich ist (Scan-Verfahren im Lager, automatische Updates von Postleitzahlen) ist diese fehlerresistenter als die manuelle Erfassung und daher im Sinne der Datenqualität vorrangig zu wählen.

Soweit bereits Prozesse bestehen, die eine Datenerhebung beinhalten (z.B. Inventur), können diese fachbereichsübergreifend um eine Datenbereinigungs-/-validierungskomponente erweitert werden.

Im Sinne eines Qualitätszirkels sind die Prozesse und Verfahren weiterzuverbessern (vgl. auch 3.6). Hier liegt die grundlegende Verantwortung zwar bei den Fachbereichen, jedoch ist die Implementierung von regelmäßigen Überprüfungen und der Umsetzung von Konsequenzen im datenbanktechnischen Bereich im Verantwortungsbereich der Datenbankverwaltung. Dies gilt insbesondere für die fortlaufende Messung der Datenqualität (s. 3.5).

## Messung und Monitoring

Die Überwachung und Messung der Datenqualität ist wichtig für jedes Unternehmen, das Entscheidungen auf der Grundlage dieser Daten trifft. Für die Messung und Überwachung der Datenqualität eignen sich verschiedene Indikatoren:  
  
Überprüfung der Daten auf Fehler oder Ungenauigkeiten. Beispiel: Vergleich von Kundendaten in der Datenbank mit externen verlässlichen Quellen, um die Richtigkeit zu gewährleisten.

Vollständigkeit: Sicherstellung, dass keine Daten fehlen, z.B. Beispiel durch Überprüfung von Datensätzen auf fehlende Werte in wichtigen Feldern wie Adressen oder Telefonnummern, Rücksprache mit den Tierpflegern, ob alle Tiere aufgenommen sind. Die Pfleger erhalten auch Zugriff für Datenabfragen.  
  
Konsistenz: Überprüfung, ob die Daten über verschiedene Systeme und Datenbanken hinweg übereinstimmen. Beispiel: Vergleich von Mitarbeiterdaten über verschiedene Systeme (z.B. Personalverwaltung) hinweg, um Diskrepanzen zu identifizieren.

Weitere Metriken sind beispielsweise Aktualität (wie aktuell die Daten sind), Einzigartigkeit (Vermeidung von Duplikaten) und Validität (ob die Daten den erwarteten Formaten und Regeln entsprechen). Dies kann z.B. durch Stichproben überprüft werden.  
  
Ein weiterer wichtiger Schritt in der Überwachung der Datenqualität ist die Implementierung eines regelmäßigen Audit-Plans. Regelmäßige Audits, ob intern oder durch externe Prüfer, sind entscheidend, um die Einhaltung der Datenstandards zu überwachen und kontinuierlich zu verbessern. Diese Audits könnten beispielsweise die Überprüfung von Datenprotokollen, die Einhaltung von Datenschutzbestimmungen und die Genauigkeit der Tierbestandsaufzeichnungen umfassen.

Die Erstellung von Qualitätsberichten ist ebenfalls ein wesentlicher Bestandteil des Prozesses. Diese Berichte könnten auf vierteljährlicher oder jährlicher Basis erstellt werden und wichtige Metriken wie die Anzahl der festgestellten Datenfehler, Probleme mit der Datenkonsistenz oder Fortschritte bei der Datenaktualisierung umfassen. Sie bieten eine Grundlage für die Bewertung der aktuellen Datenqualität und für die Planung von Verbesserungsmaßnahmen und sollen der Geschäftsleitung vorgelegt werden.

## Problemmanagement und Eskalationsverfahren

Für die Behebung einzelner kleinerer Fehler ist in der Regel die Fachabteilung zuständig. Soweit die Datenbankkonzeption ein häufigeres Auftreten bestimmter Fehler nicht verhindert, ist dann im Benehmen mit der Datenbankadministration fachbereichsübergreifend ein Projekt zur Vermeidung dieser Fehler zu konzeptionieren und umzusetzen. Die Verantwortung hierfür liegt bei den Fachbereichen.

Die Feststellung von größeren Problemen bei der Datenqualität kann im Rahmen eines regelmäßigen Monitorings oder anlassbezogen (z.B. starke Inventurabweichungen) auftreten. Gewichtige Probleme sind der Geschäftsleitung mitzuteilen. Im Sinne eines Risikomanagements sind daraufhin durch die Geschäftsleitung – gegebenenfalls unter Einbeziehung Dritter – geeignete Maßnahmen zur Behebung zu projektieren und die entsprechenden Ressourcen bereit zu stellen.

## Schulung und Bewusstseinsbildung

Bei Einführung der digitalen Verwaltung sind Schulungen aller Mitarbeitenden zu den Grundzügen des Umfangs der Datenbank und der Möglichkeiten der Nutzung notwendig. Für die Datenqualität relevant sind insbesondere die eingabebefugten Mitarbeitenden, denen die Vorgaben der Datenstruktur in der Datenbank selbst, die Bedeutung der einzelnen Felder sowie die zugehörige Erläuterung im „data dictonary“ ausführlich darzustellen ist und die intensivere Schulungen und Trainings erhalten sollten.

Hierfür eignet sich eine „Schulungsdatenbank“, in der die Mitarbeitenden ohne Auswirkungen auf die „echte“ Datenbank Eingaben vornehmen und testen können. Hier können auch Trainingsbeispiele mit den entsprechenden Auswirkungen auf Reports angelegt werden, um die Folgen schlechter Datenqualität darzulegen (z.B. Doppelzählungen von Tieren, fehlende Futterbestellungen, etc.).

Für diese Schulungen ist ein nutzernahes, den Prozessabläufen folgendes (digitales) Handbuch zu erstellen, das mit den Nutzern gemeinsam während der Schulungen erweitert und ergänzt werden sollte, um deren Erfahrung zu nutzen. Eine laufende Aktualisierung des Handbuchs ist dabei vorzusehen.

Konzeptioniert ist das Handbuch zur Nutzung als Nachschlagewerk für die Mitarbeitenden sowie als Schulungsleitfaden für neue Mitarbeitende.

Bei Ersterfassung der Altdaten und bei der Erfassung von Daten durch neue Mitarbeitende ist die Umsetzung des 4-Augen-Prinzips empfehlenswert.

## Kontinuierliche Verbesserung

Die kontinuierliche Verbesserung der Datenqualität im digitalisierten Zoo kann durch die Einrichtung eines Mechanismus für die regelmäßige Bewertung der Daten erfolgen. Dies kann z.B. durch die Implementierung eines Dashboards erfolgen, das Echtzeit-Indikatoren zur Datenqualität wie Genauigkeit, Vollständigkeit, Konsistenz und Aktualität der Informationen anzeigt. Dies ermöglicht es den Mitarbeitern, Probleme schnell zu identifizieren und anzugehen.

Regelmäßige Daten-Reviews und Audits sollen Teil der Routine des Zoos sein, wobei Datenqualitätsberichte in periodischen Meetings analysiert und besprochen werden. Externe Audits können dabei helfen, eine objektive Bewertung der Datenqualität sicherzustellen. Ein weiterer wichtiger Aspekt ist die Einrichtung von Feedback-Schleifen. Mitarbeiter, die täglich mit den Daten arbeiten, sollten ein einfaches Verfahren haben, um Rückmeldungen zur Datenqualität zu geben, sei es durch ein internes Ticket-System oder regelmäßige Umfragen.

Die Anpassung der Datenqualitätsstrategie basierend auf dem gesammelten Feedback und neuen Erkenntnissen ist dabei entscheidend. Dies kann die Einführung neuer Technologien zur Datenverarbeitung, die Schulung der Mitarbeiter in neuen Methoden der Datenerfassung oder die Anpassung der bestehenden Datenstrukturen beinhalten.

Dabei könnte auch der Einsatz von modernen Datenanalysetools, einschließlich maschinellem Lernen und künstlicher Intelligenz, dabei helfen, Muster und Trends in den Daten zu identifizieren, die auf Probleme mit der Datenqualität hinweisen könnten. Durch diese Maßnahmen kann der Zoo eine Kultur der kontinuierlichen Verbesserung schaffen, die es ihm ermöglicht, seine Datenqualität ständig zu überwachen und zu verbessern.  
  
Die Vernetzung mit anderen Zoos wäre ein weiterer wichtiger Schritt. Durch den Austausch von Informationen können Zoos von gemeinsamen Erkenntnissen profitieren. Eine solche Kollaboration ermöglicht z.B. die Entwicklung gemeinsamer Standards für die Datenqualität, wodurch eine einheitliche Datenqualität über verschiedene Einrichtungen hinweg sichergestellt wird. Zusätzlich bietet der Austausch mit anderen Zoos die Möglichkeit, Best Practices im Datenmanagement zu identifizieren und durch Benchmarking voneinander zu lernen. Gemeinsame Forschungs- und Entwicklungsprojekte können zudem neue Methoden zur Datenerfassung und -analyse hervorbringen. Der regelmäßige Austausch und das Feedback zwischen den Zoos unterstützen die kontinuierliche Verbesserung der eigenen Datenqualitätsstrategien.

## Compliance und rechtliche Anforderungen

Die laufende Sicherstellung datenschutzrechtlicher Vorgaben ist durch die Benutzerrechteverwaltung, die räumliche Sicherung und die weiteren Sicherheitsmaßnahmen (s. Sicherheitskonzept) gegeben.

Die fachliche Prüfung und Sicherstellung fachlicher rechtlicher Anforderungen (z.B. Tierschutz) obliegt den jeweiligen Fachbereichen, gegebenenfalls mit Unterstützung externer Dritter. Dies gilt insbesondere für Änderungen im rechtlichen Bereich. Die Umsetzung in der Datenbank wird durch den Datenbankadministrator vorgenommen, in der Regel ist hierzu eine fachbereichsübergreifende Projektarbeit notwendig.

# Zusammenfassung

Das Ziel der Digitalsierung des Zoos in Pirmasens zielt auf die Entwicklung eines integrierten IT-Systems zur Modernisierung der Zoo-Verwaltung ab. Der Schwerpunkt liegt auf der Digitalisierung und Verbesserung von Prozessen, insbesondere in der Erfassung und Verwaltung von Daten über Tiere, Gehege, Mitarbeiter und Futterlieferungen. Dazu gehört die Entwicklung einer relationalen Datenbank, eines Data Warehouse und das Erreichen hoher Datenqualität.

Im Rahmen dieses Projekts werden detaillierte Informationen zu Tierpflegern, Tieren, Gehegen, Tierärzten und Futterlieferanten erfasst. Der Zoo soll vergrößert werden und plant auch die Erweiterung seiner IT-Infrastruktur, die Einführung eines Webshops und virtueller Rundgänge. Business Intelligence-Analysen sollen vielfältige Dimensionen wie Gattungen, Krankheitsverläufe, Mitarbeiterdaten und mehr abdecken. Wichtige Geschäftsvorfälle beinhalten die Anlage neuer Tiere, Gehege, Mitarbeitermanagement, Fütterungs- und Bestandsverwaltung, Rundgänge, Krankenakten und die Verwaltung von Tierarztinformationen.

Eine besondere Herausforderung stellt die Digitalisierung einer bislang analog geführten Datenhaltung und die fehlende IT-Erfahrung der Mitarbeiter dar.

Entwickelt wurde eine benutzerfreundlichen Datenbank. Ein halbautomatisches Verfahren überprüft die Konsistenz der Datenbank mit dem erstellten Datenmodell. Ein Data Dictionary unterstützt die Mitarbeiter bei der korrekten Dateneingabe und -pflege.  
  
Referenztabellen für bestimmte Wertebereiche und Konsistenzprüfungen wurden eingeführt. Die Planung der Datenerfassung und regelmäßige Qualitätskontrollen sollen Risiken minimieren. Data Profiling nach der Ersterfassung prüft auf Fehler wie fehlende Werte und Doppelerfassungen.

Die klare Dokumentation des Modells und die Schulung der Mitarbeiter unterstützen die Akzeptanz und Umsetzung der Digitalisierung. Benutzeroberflächen für konsistente Dateneingaben und die Definition von Testfällen für typische Geschäftsvorfälle wurden ebenfalls umgesetzt.

Die Sicherstellung der Datenqualität im laufenden Betrieb des Zoo Pirmasens erfordert umfassende Maßnahmen, um den Herausforderungen neuer Datenerfassungen, der Datenalterung und möglicher Gefahren durch fehlerhafte oder fehlende Daten zu begegnen. Es ist entscheidend, Inkonsistenzen und Fehler im Datenbestand zu vermeiden, die zu Gefährdungen durch oder für Tiere führen könnten. Ebenso wichtig ist die Prävention von Datenverlust durch technische Ausfälle, Fehlbedienungen oder Kriminalität, sowie die strikte Einhaltung aller relevanten Datenschutzgesetze und -vorschriften.  
  
Die Ziele der Datenqualität umfassen eine angestrebte Datengenauigkeit von 97%, wobei spezifische Anforderungen je nach Relevanz der Datenbereiche variieren. Dies betrifft insbesondere sensible Bereiche wie die Verwaltung der Tiere, wo eine höhere Datenqualität erwartet wird. Um die Konsistenz und korrekte Weiterverarbeitung von Daten zu gewährleisten, sind unternehmensinterne Richtlinien zur Datenerfassung, wie die einheitliche Erfassung von Daten, notwendig. Diese Richtlinien sollen regelmäßig überprüft und erweitert werden, und alle erfassenden Mitarbeiter müssen Zugang zu ihnen haben. Wertelisten für bestimmte Datenbereiche sollen Fehleingaben verhindern und können in anderen Bereichen ergänzt werden.  
  
Die Verantwortung für die Datenverfügbarkeit und -qualität liegt bei der Geschäftsleitung, während die Abteilungen und Endbenutzer in die Verantwortung für die enthaltenen Daten einbezogen werden. Eine nutzungsorientierte Rechteverwaltung ist für den Datenschutz und kritische Datenbereiche erforderlich, mit einer restriktiven Benutzerrechtsvergabe basierend auf fachlicher Kompetenz und Verantwortung.

Die Prozesse und Verfahren bei der Datenerfassung und Fehlerbehandlung sind klar zu definieren und schriftlich festzuhalten. Einheitliche Vorgehensweisen und der einfache Zugang zu grundlegenden Dokumenten wie dem Data Dictionary sind für alle Endnutzer zu ermöglichen. Unterstützte Erfassungsverfahren sind aufgrund ihrer geringeren Fehleranfälligkeit bevorzugt einzusetzen.

Die Überwachung und Messung der Datenqualität erfolgt durch verschiedene Indikatoren wie Fehlerprüfung, Vollständigkeit und Konsistenz. Regelmäßige Audits und die Erstellung von Qualitätsberichten sind entscheidend, um die Einhaltung der Datenstandards zu überwachen und kontinuierlich zu verbessern.

Im Bereich Schulung und Bewusstseinsbildung sind umfassende Schulungen für alle Mitarbeiter erforderlich. Intensivere Trainings sind für dateneingabebefugte Mitarbeiter vorgesehen, unterstützt durch eine Schulungsdatenbank für praktische Übungen.

Die kontinuierliche Verbesserung der Datenqualität wird durch regelmäßige Bewertungen, Daten-Reviews, externe Audits und Feedback-Schleifen für Mitarbeiter gefördert. Die Anpassung der Datenqualitätsstrategie basierend auf gesammeltem Feedback und neuen Erkenntnissen ist entscheidend.

Die Vernetzung mit anderen Zoos bietet die Möglichkeit, Informationen auszutauschen, gemeinsame Standards für Datenqualität zu entwickeln und voneinander zu lernen. Die fachliche Prüfung und Sicherstellung rechtlicher Anforderungen obliegt den jeweiligen Fachbereichen, unterstützt durch externe Dritte, und ist für die Umsetzung rechtlicher Änderungen in der Datenbank notwendig.

Bei der Erstellung eines Datenqualitätsplans oder für Schulungen stehen wir sehr gerne unterstützend zur Verfügung.

Anhang 1 Maße zur Bewertung der Data Quality

1. **Genauigkeit**: Dieser Indikator misst, wie korrekt und fehlerfrei Daten in einem System erfasst sind. Es geht um die Übereinstimmung der Daten mit der Realität oder einer verlässlichen Quelle.
2. **Vollständigkeit**: Dieser Indikator bewertet, inwiefern alle erforderlichen Daten erfasst und in den Datensätzen verfügbar sind. Fehlende Daten können zu falschen Schlussfolgerungen oder Ineffizienzen führen.
3. **Konsistenz**: Konsistenz misst, ob dieselben Daten über verschiedene Systeme und Datenbanken hinweg gleich bleiben. Inkonsistenzen können Verwirrung stiften und die Glaubwürdigkeit der Daten beeinträchtigen.
4. **Zuverlässigkeit**: Dieser Indikator bewertet, wie vertrauenswürdig und glaubwürdig die Daten sind. Er hängt eng mit der Genauigkeit und dem Quellenmanagement zusammen.
5. **Aktualität**: Dieser Indikator gibt an, wie aktuell die Daten sind. Veraltete Daten können zu schlechten Entscheidungen führen, besonders in schnelllebigen Umgebungen.
6. **Einzigartigkeit**: Misst, inwieweit Duplikate in den Datenbeständen vermieden werden. Mehrfacheinträge derselben Daten können zu ineffizienten Prozessen und verwirrenden Analysen führen.
7. **Integrität**: Dieser Indikator misst, inwieweit die Daten in der Datenbank die realen Beziehungen und Regeln der Geschäftswelt korrekt abbilden. Zum Beispiel sollten in einer relationalen Datenbank die Beziehungen zwischen verschiedenen Tabellen logisch und konsistent sein.

Anhang 2 Allgemeine Prinzipien zur Sicherstellung der Data Quality

**Datenintegrität**: Es wird sichergestellt, dass die Daten korrekt, konsistent und verlässlich sind, indem Datenintegritätsregeln wie Primär- und Fremdschlüssel implementiert werden.

**Datenkonsistenz**: Die Konsistenz der Daten über die gesamte Datenbank hinweg wird durch konsistente Datenmodelle und Geschäftslogik gewährleistet.

**Datenvalidierung und -verifizierung**: Die Daten werden bei der Eingabe auf Genauigkeit und Vollständigkeit geprüft, indem Validierungsregeln angewendet werden.

**Datenaktualität**: Die Aktualität und Relevanz der Daten werden durch regelmäßige Aktualisierung und Wartung sichergestellt.

**Skalierbarkeit und Erweiterbarkeit**: Zukünftiges Wachstum und Erweiterungen der Datenbank werden durch eine skalierbare Datenbankstruktur ermöglicht.

**Klare Verantwortlichkeiten**: Für jeden Aufgabenbereich und Prozess werden klare Verantwortlichkeiten definiert.

**Dokumentation und Wissensmanagement**: Ausführliche Dokumentationen und das Teilen von Wissen innerhalb der Organisation tragen zum Verständnis und effektiver Nutzung der Datenbank bei.

**Datensicherheit**: Schutz der Daten vor unautorisiertem Zugriff und Beschädigung wird durch Maßnahmen wie Zugriffskontrollen und Verschlüsselung erreicht.

**Backup und Disaster Recovery**: Backup-Strategien und Disaster Recovery-Pläne werden etabliert, um Datenverlust bei Systemausfällen zu vermeiden.

**Performance-Management**: Die Performance der Datenbank wird durch effizientes Design, Indizierung und Abfrageoptimierung verbessert.

**Compliance und rechtliche Anforderungen**: Die Einhaltung von Gesetzen und Compliance-Anforderungen wird durch regelmäßige Überprüfungen und Anpassungen der Datenbank sichergestellt.

Anhang 3 Erstellung eines Datenqualitätsplans

Der Datenqualitätsplan ist ein strategisches Dokument, das Richtlinien und Maßnahmen zur Sicherstellung, Messung und Aufrechterhaltung der Qualität von Daten innerhalb des Unternehmens und/oder einer spezifischen Datenbank definiert. Der Plan umfasst folgende Kernaspekte:

1. Ziele der Datenqualität:

* Definition spezifischer, messbarer Ziele, die erreicht werden sollen, wie z.B. die Genauigkeit, Vollständigkeit, Konsistenz und Zuverlässigkeit von Daten.

2. Standards und Richtlinien:

* Festlegung von Standards für die Datenerfassung, -bearbeitung und -speicherung.
* Richtlinien für Dateneingabe, Formatierungsregeln, Validierungsverfahren und Umgang mit Daten.

3. Verantwortlichkeiten:

* Zuweisung von Verantwortlichkeiten für verschiedene Aspekte der Datenqualität, einschließlich der Rollen und Aufgaben von Datenstewards, IT-Personal und Endbenutzern.

4. Prozesse und Verfahren:

* Detaillierte Beschreibung der Prozesse zur Datenerfassung, Datenbereinigung, Datenvalidierung und -verarbeitung.
* Verfahren zur regelmäßigen Überprüfung und Aktualisierung von Daten.

5. Messung und Monitoring:

* Methoden zur Überwachung und Messung der Datenqualität, einschließlich der Verwendung von Key Performance Indicators (KPIs) und anderen Metriken.
* Plan für regelmäßige Audits und Qualitätsberichte.

6. Problemmanagement und Eskalationsverfahren:

* Prozess zur Identifizierung, Meldung und Behebung von Datenqualitätsproblemen.
* Eskalationsverfahren für schwerwiegende oder systematische Datenqualitätsprobleme.

7. Schulung und Bewusstseinsbildung:

* Programme zur Schulung der Mitarbeiter in Bezug auf die Bedeutung der Datenqualität und die korrekte Handhabung von Daten.
* Sensibilisierungskampagnen, um die Wichtigkeit guter Datenqualität zu betonen.

8. Kontinuierliche Verbesserung:

* Vorgehensweise für die kontinuierliche Bewertung und Verbesserung der Datenqualität.
* Anpassung der Datenqualitätsstrategie basierend auf Feedback und neuen Erkenntnissen.

9. Compliance und rechtliche Anforderungen:

* Berücksichtigung rechtlicher und regulatorischer Anforderungen bezüglich der Datenspeicherung und -verarbeitung.

Ein Datenqualitätsplan nach diesem Muster ist vom Datenverantwortlichen Mitarbeiter/Stab zu erstellen