

**Auftraggeber:** City Zoo

Ansprechpartner: Herr Teske

**Projekt:** IT-gestützter Zoobetrieb

Konzept für Datenqualität (Version 05.01.2024)

**Auftragnehmer:** Müller AG

Gruppe 02

Bartsch, Grapengeter, Steinle, Mehmeti, Bezzubov

## Inhaltsverzeichnis

Projektvorstellung .....	2
Aktuelle Situation .....	2
Geplante Veränderungen .....	2
Technische Lösung .....	3
Datenqualität: Probleme und Lösungen.....	3
Fehlerarten, Ursprung .....	4
Zusammenfassende Anmerkungen:.....	6
Überblick über wichtigste Gegenmaßnahmen:.....	7
Weitere Schritte.....	7

## Projektvorstellung

### Aktuelle Situation

Alle Unterlagen im Tiergarten Pirmasens werden seit 36 Jahren dezentral erstellt und verwahrt. Im Tiergarten kümmern sich 70 Mitarbeiter um 6000 Tiere. Tiere sind Tierarten zugeordnet. Von einzelnen Tierarten befinden sich mehrere Exemplare im Tiergarten. Der Tiergarten unterhält internationale Geschäftsbeziehungen zu 50 Tierärzten aus unterschiedlichen Praxen und zu 120 Futterlieferanten.

### Geplante Veränderungen

Künftig sollen alltägliche Geschäftsprozesse/Ereignisse des Tiergartens auf IT-gestützte Datenerfassung und Datenauswertung umgestellt. Unter anderem:

- Aufnahme eines neuen Tieres / in diesem Zuge ggf. auch neue Gattung / Tierart anlegen
- Unterbringung und Verlegung der Tiere unter Berücksichtigung der Unverträglichkeiten
- Geburt von Jungtieren und deren Abstammung
- Krankheit bei Tieren / Krankheitsverlauf
- Anlage eines neuen Arztes
- Zuständigkeiten der Ärzte für Tiere und deren Veränderungen
- Behandlung durch Ärzte (akut und nach Plan)
- Ärztevertretung im Krankheitsfall
- Ärztevertretung bei Urlaub
- Beitritt/Austritt von Mitarbeitern
- Zuständigkeiten von Mitarbeitern für Einzeltiere und deren Veränderungen
- Mitarbeitervertretung im Krankheitsfall (Planung und Erfassung)
- Mitarbeitervertretung bei Urlaub (Planung und Erfassung)
- Fütterung der Tiere (Zubereitung und Durchführung)
- Futterbestellung
- Lagerung Futter, Bestandsverwaltung
- Anlage eines neuen Futterlieferanten
- Planung eines Rundganges (in Person)

Künftig müssen auch Recherchen über folgende Bereiche möglich sein:

- Gattung: z.B. welche sind im Zoo vertreten
- Tierart: z.B: welche sind verwandt
- Tier: z. B: wo befindet sich momentan das Tier, wann/womit war es gefüttert
- Krankheit / Krankheitsverlauf: Ursachen, Anzahl der kranken Tiere
- Mitarbeiter: Dienst, Zuständigkeiten
- Gehege / Gehegebelegung: Zustand, Belegung
- Futterlieferant: Futterarten, Rabatte
- Futterart: Lieferanten, Rabatte
- Futter: verbrauchende Tiere
- Futterverbrauch
- Lager / Lagerbestand
- Tierarzt
- Zeit
- Rundweg

Nach der Umstellung sollen alle Mitarbeiter auf einen konsistenten, validen und aktuellen Datenbestand zugreifen können. Als Zielwert für die Datenqualität sind >97% vorgegeben. Die Priorität liegt auf Aktualität und Richtigkeit der Daten.

Die einzuführenden IT-Datentools müssen eine Erweiterung der Tier-, Lieferanten- Gebäude- und Mitarbeiterbestände erfassen können. Außerdem werden künftig virtuelle Rundgänge durch den Zoo angeboten.

## Technische Lösung

Für Abbildung des operativen Geschäftes wird eine rationale Datenbank (auf Access-Basis) und für analytische Zwecke ein Datawarehouse auf Basis vom Data-Vault-Modell eingeführt.

Die aktuellen analogen Daten (ALTDATEN) werden in das neue digitale System durch einen externen Dienstleister überführt. Die Planbarkeit, Rückverfolgbarkeit und leichte Ausbaufähigkeit des Systems sind von oberster Priorität, da der Kunde weitere Ausbauten des Betriebs in Form von neuen Gebäuden, Tieren und Mitarbeitern plant. Wichtige Erweiterungen in der Zukunft sind außerdem die Möglichkeit von Tierpatenschaften, einem Webshop und virtuellen Führungen.

## Datenqualität: Probleme und Lösungen

Die Datenqualität hat Einfluss auf kritische und weniger kritische Bereiche in der Tätigkeit eines Zoos. Zu den kritischen Bereichen gehören die Gesundheit der Tiere, Sicherheit der Besucher und der Mitarbeiter. Zu weniger kritischen Bereichen gehören betriebswirtschaftliche Auswirkungen durch eine unzureichende Datenqualität.

Datenqualität von >97% bedeutet, dass in 10000 Datensätzen nicht mehr als 299 Fehler vorkommen dürfen.

Kriterien der Datenqualität sind Vollständigkeit, Eindeutigkeit, **Korrektheit**, **Aktualität**, Genauigkeit, Konsistenz, Redundanzfreiheit, Relevanz, Einheitlichkeit, **Zuverlässigkeit** und Verständlichkeit.

## Fehlerarten, Ursprung

(A) Wir unterscheiden zwischen den Fehlern durch Eingabe in operative Datenbank und durch Übertragung aus dieser Datenbank ins Datawarehaus.

(B) Aus Perspektive der operativen Datenbank kann speziell zwischen folgenden Fehlern unterschieden werden:

- (1) Fehler aus der „analogen“ Vergangenheit (Altdaten), (2) Fehler aus künftiger Fehlbedienung der Datenbank (Neudaten) und (3) Fehler aus falscher Konstruktion der Datenbank (Datenformate, Verknüpfungen)
- (2) Fehler in den Altdaten können lediglich kontrolliert und ausgebessert werden (Scherpunkt *Kontrolle*).
- (3) Fehler in den Neudaten können im laufenden Betrieb entweder vorgebeugt oder mindestens stark eingeschränkt werden (Schwerpunkt *Fehlervermeidung*).
- (4) Fehler bzw. Datenkonflikte durch falsche Konstruktion der Datenbanken können nur bedingt beim laufenden Betrieb der Datenbanken ausgebessert werden und i.d.R. ohne positive Auswirkung auf bereits gespeicherte Daten (Schwerpunkt *Fehlervermeidung*)

(C) Aus inhaltlicher Perspektive kann es zwischen folgenden Fehlern unterschieden werden:

Werte **veraltet**, Werte **fehlen**, Werte **unvollständig**, Werte **falsch**, Werte **doppelt**, Werte **widersprüchlich**

Mögliche Problemfelder bei operativen Daten:

(1) **Angaben sind veraltet** (mögliche Beispiele: Adresse, Familienname, Lieferantename, Futterbezeichnung stimmen nicht mehr).

### Ursache:

viele Informationen wurden erst vor mehreren Jahren erfasst (z.B.: bei Aufnahme eines neuen Mitarbeiters/Lieferanten) und nicht mehr angefasst. Nicht alle Veränderungen (Kontaktdaten, Namen, FiBu-Daten) waren evtl. später erfasst. Für diese Fehlerart wären insbesondere ALTDATEN (aus dem bisherigen analogen System) anfällig.

### Gegenmaßnahmen:

(A) bei kleineren Datenmengen (wie Personal) – Prüfung der Daten durch Mitarbeiter selbst (1:1) / Kontrollliste (Mitarbeiter /Daten geprüft)

(B) bei mittleren Datenmengen wie Lieferanten/Ärzte :

(a) alle Geschäftsverbindungen werden gebeten, Daten zu überprüfen

(b) Abgleich von Daten mit ersten neuen Lieferscheinen / Rechnungen durch Mitarbeiter innerhalb eines vordefinierten Zeitraums, anschließend eine regelmäßige Überprüfung

(analog zur Qualitätskontrolle bei Neudaten), um selten genutzte Geschäftsverbindungen in die Kontrolle miteinzubeziehen.

(C) bei größeren Datenmengen wie Tiere: verstärkte Überprüfungsmaßnahmen in den ersten Tagen nach Inbetriebnahme der Datenbank gepaart mit anstehenden Aktionen (z.B.: Selektion von Tieren für Gesundheits-Check nach bestimmten Arten oder nach Datum der letzten Untersuchung).

(2) **Angaben unvollständig** (mögliche Beispiele: Messeinheiten fehlen, Adresse nur teilweise vorhanden, bekannte Beispiele; Angaben in der Referenzliste PLZ/Orte).

Ursache: Einige Informationen wurden evtl. ohne Anlehnung auf bestimmte Strukturen oder Formatvorgaben erfasst. Daraus resultieren Fehleroptionen.

Gegenmaßnahmen: maximal mögliche Aufteilung von Daten; laufende Kontrolle von Datenfehlern.

(3) **Angaben falsch** (Beispiele: o.g. und z.B. Feldnamen in den Tabellen mit ALTDATEN).

Ursachen: Schreibfehler; nicht übereinstimmende Feldnamen in Quelltabellen und Empfänger-Tabellen, falsch gewählte Maßstäbe bei Messwerten

Gegenmaßnahmen:

(A) bei numerischen Werten Plausibilitätsprüfung

- a. statistische Untersuchung (gegen Mittelwert und SV), gegen Sollwert-Spannen (sind noch nicht bekannt)
- b. „weiche“ Einschränkung der Eingabe durch hinterlegte Wertspannen (kein Ausschluss, aber Warnung).

(B) bei fehlerhaften Textdaten (wie Tiernamen, Tierart):

- a. müssen die zugeordneten Tierpfleger die Daten prüfen (auf Aufforderung bei täglicher Arbeit in zeitlichen Intervallen)
- b. Eingabe über vordefinierte Werte (Dropdown-Minus),

(4) **Mehrfache Erfassung** (Beispiele: mehrere identische Datensätze für ein und dasselbe Tier sind vorhanden; mehrere leicht abweichende Datensätze (z. B.: Angaben zu Eltern/Geburtsdatum/Herkunft) für ein und dasselbe Tier sind vorhanden).

Ursache: fehlende Kontrolle der wiederholten Eingabe

Gegenmaßnahmen: Häufigkeiten der Datensätze zählen und Abgleich mit Erwartungswerten (6000 Tiere, 70 Mitarbeiter, 140 Futterlieferanten, Gehege/Haus-Anzahl, 50 Ärzte) berechnen (wenn = 0, dann kein Handlungsbedarf) ODER Tools zu Dubletten-Bereinigung einsetzen. Worthäufigkeiten in Datenfeldern zählen und mit Erwartungswerten (z.B. Fütterungen) abgleichen (davor werden die Datensätze maximal normalisiert/atomisiert)

(5) **Altdaten/Neudaten sind aktuell, richtig und vollständig, aber nicht geeignet**

**strukturiert/formatiert:** und dahernicht direkt in operative Datenbank/DWH einlesbar: z.B. unterschiedliche Trennzeichen in den tabellarisch strukturierten txt-Dateien. Für diese Fehlerart wären insbesondere ALTDATEN (aus dem bisherigen analogen System) anfällig.

Ursache: fehlende Vorgaben für Datenformate bzw. fehlende Kontrollmechanismen

Gegenmaßnahmen: Vorgaben für Datenformate bzw. Kontrollmechanismen

(6) **Unnötige** (weniger relevante) **Daten**: Zeitverlust bei Erfassung.

Ursache: fehlende Priorisierung und Zweckbindung der erhobenen Daten

Gegenmaßnahmen: Nutzen der zu erhebenden Daten und den Aufwand abwägen

## Zusammenfassende Anmerkungen

1. Je mehr Felder als MUSS-Felder definiert werden, je genauer Datenformate beschrieben werden, je mehr Eingaben durch Auswahlmenüs unterstützt werden, desto weniger Platz bleibt für fehlerhafte Datenpflege.
2. Wegen ungleichmäßiger Belastungsverteilung (Hochsaison, zeitweise Personalengpässe) muss aber für jedes Geschäftsprozess zwischen kritischen und optionalen Datenfeldern unterschieden werden. Es wird empfohlen auf Basis der *Szenarienmodellierung (Was-Wenn)* solche aufzuspüren.
  1. **Fehlende kritische Daten** (*müssen auch vordefiniert werden*)
    - (A) **Fehlende Einträge in Datensätzen** bei Tier, Gehege, Pfleger, Arzt, Futterbestand, Lieferant, Rundweg(Abschnitt); **Folge**: wenn nicht erfasst, Funktion nicht möglich >>keine Fütterung, kein Gesundheitscheck, keine Tierpflege, **Gegenmittel**: MUSS-Felder definieren
    - (B) **fehlende Datensätze** wie Krankheits-/Behandlungseintrag; Gesundheitsrisiken für Tiere, keine notwendige Sicherheit (mehrere Tierpfleger bei bestimmten Tierarten) u.s.w. **Gegenmaßnahmen**: Fällige Aktionen nur nach Erledigen von Datenpflege: z.B. Keine Rechnungsfreigabe ohne Eintrag in die Tierakte und ohne Befund (auch über zu befüllende Felder in den Datensätzen für fällige Aktionen steuerbar)
  2. **Fehlende optionale Daten**: optische Merkmale eines Tieres, Geburtstag eines Tierpflegers, wenn nicht erfasst, evtl. falsche Auswertung oder eine noch nicht angedachte Funktion künftig nicht möglich. Qualitätssicherungsmaßnahmen: **Gegenmittel**: Leere Felder in gewissen zeitlichen Abständen zählen und z.B. zugeordnete Pfleger ansprechen auf Datenvervollständigung
  3. **Falsche Daten**: falsche Futterzusammensetzung, falsche Medikation, nicht rechtzeitige Fütterung/Gesundheitscheck, falsche Fütterung, falsche Größen (wenn falsche Einheiten kg statt gr, m<sup>3</sup> statt m<sup>2</sup> gewählt wird). **Gegenmittel**: Plausibilitätsprüfung (Abgleich mit Wertspannen)
  4. **Widersprüchliche Daten**: ausgeschlossen, wenn keine Datenredundanz vorkommt und Datenintegrität gewährleistet wird

## Überblick über wichtigste Einzelmaßnahmen

- a. Vordefinierte Eingabe / Dropdown-Menüs (Folge: Einheitlichkeit und keine Rechtsschreibfehler)

- b. MUSS-Felder, Zählen von gefüllten KANN-Feldern und Abgleich mit erwarteten Zielwerten (Folge: kritische Angaben immer vorhanden, fehlende optionale Angaben leicht auffindbar)
- c. für migrierte Altdaten: Datenüberprüfung / statistische Kontrollen (z.B. Plausibilitätsprüfung der Gewichtsangaben: Wertspannen zum Abgleich, Ausreißer bei Eingabe abfangen)
- d. für Angaben über operative Datenbank: hinterlegte Messeinheiten und „weiche“ Wertspannen für Größe-/Gewichteingaben bei gegebenen Tierangaben (z.B.: wenn Gehege ein Aquarium ist, dann Volumen statt Fläche; wenn bestimmte Art der Tiere aus eine Liste, dann hinterlegte Einheiten)
- e. in die Nutzeroberfläche eingebaute oder manuell gesteuerte Aufforderungen, die Daten (vor einer Aktion) zu aktualisieren.

## Weitere Schritte

- Altdaten vor Eingabe in tabellarischer Form zusammenfassen und maximale Datenlängen (Namen, Rufnummern) messen und ggf. Feldgrößen in Datenbank-Tabellen anpassen (um Abschneiden von Teildaten bei Migration zu vermeiden)
- Notwendige Felder auf „Eingabe erforderlich“ („keine NULL-Werte zulässig“) vor dem Erststart der operativen Datenbank durch den Tiergarten umstellen (*Datenbankprototyp anpassen*)
- Eingabe-Interface für operative Datenbank muss zeitgleich mit dem Start der operativen Datenbank in Betrieb genommen werden. Falls das Frontend mit Access-Mitteln umgesetzt wird, können im die Lieferumfang enthaltenen Referenztabellen als Auswahllisten für Eingabe (Dropdown-Menüs) herangezogen werden. Es reduziert die Eingabezeit und die Fehlerquote durch das Vertippen.