Projekt Datenmodell Zoo Pirmasens

Wegmann AG
Tobias Berger, Mohammad Hessan, Jerg Jaisle, Vsevolod Dorskiy, Tobias Gründer

Agenda

- Anforderungen und Ziele
- ERM
- Operative Datenbank
- Data Dictionary
- DWH: Data Vault
- Tools
- Data Quality
- Nächste Schritte
- Fragen

Anforderungen und Ziele

- Digitalisierung der Geschäftsprozesse
- Migration der analogen Daten (nur aktuell)
- Erstellung eines Operativsystems (Datenbank) und einer IT-Architektur für ein Datawarehouse für spätere BI-Analysen, sowie ein Konzept zur Datenqualität

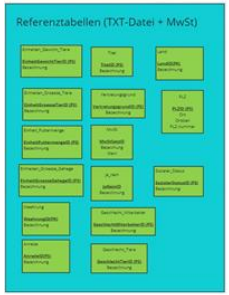
Anforderungen und Ziele

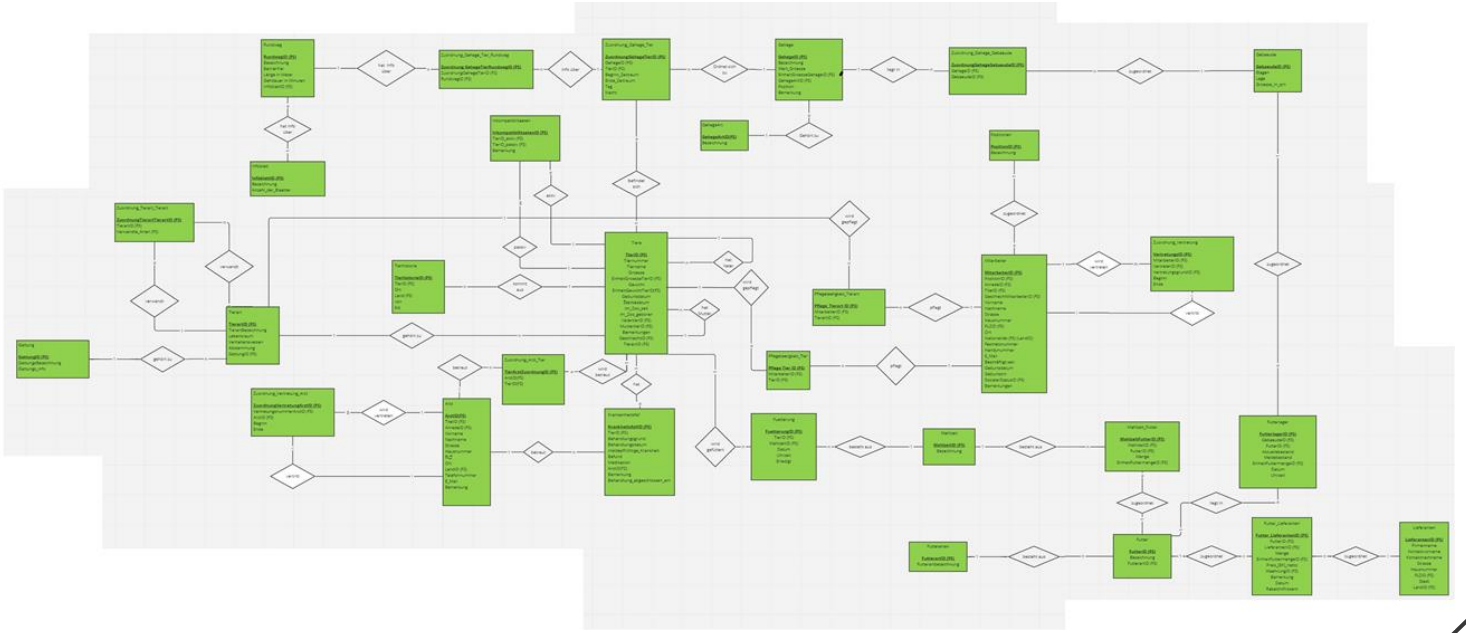
- Kick-Off Gespräch mit Frank Teske hat stattgefunden und wurde dokumentiert
- Gruppe 03 hat verschiedene Fragen gestellt, die beantwortet wurden
  - z.B. Detaillevel Abbildung der Gebäude, Datenqualität
- Kunde hatte im Rahmen des Gesprächs noch wertvolle Zusatzinfos
  - Die wichtigsten Geschäftsvorfälle
  - BI-Dimensionen für spätere Analysen

Anforderungen und Ziele

- Stakeholder
  - Kunde/Auftraggeber (Zoo Pirmasens, Ansprechpartner Herr Teske)
  - Mitarbeiter des Kunden (Benutzer des IT-Systems)
  - Auftragnehmer (Firma Wegmann AG)
- Ziele des Projekts
  - Operativsystem (Datenbank) und Datenmodell Datawarehouse für Kunde
- Kunde führt die initiale Befüllung des Operativsystems selbst durch
- Kunde schafft zwei IT-Stellen für Implementierung und Maintenance

ERM

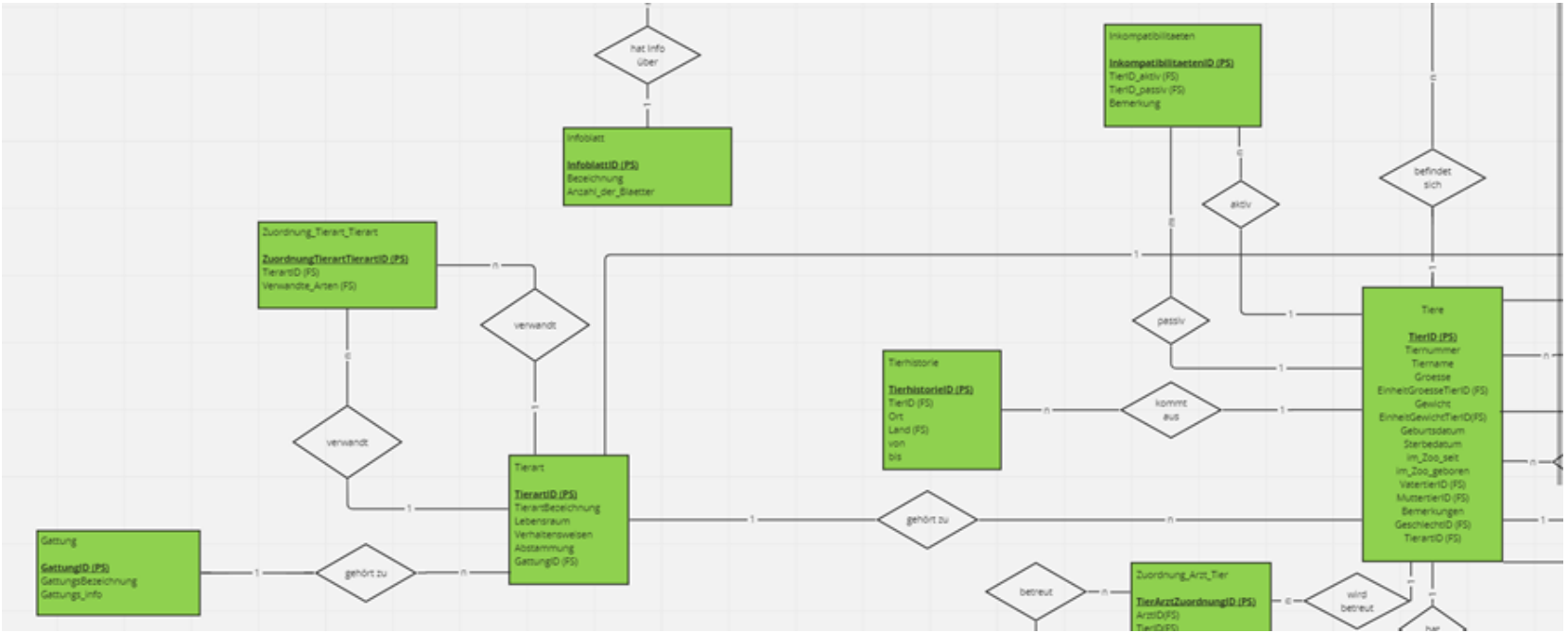




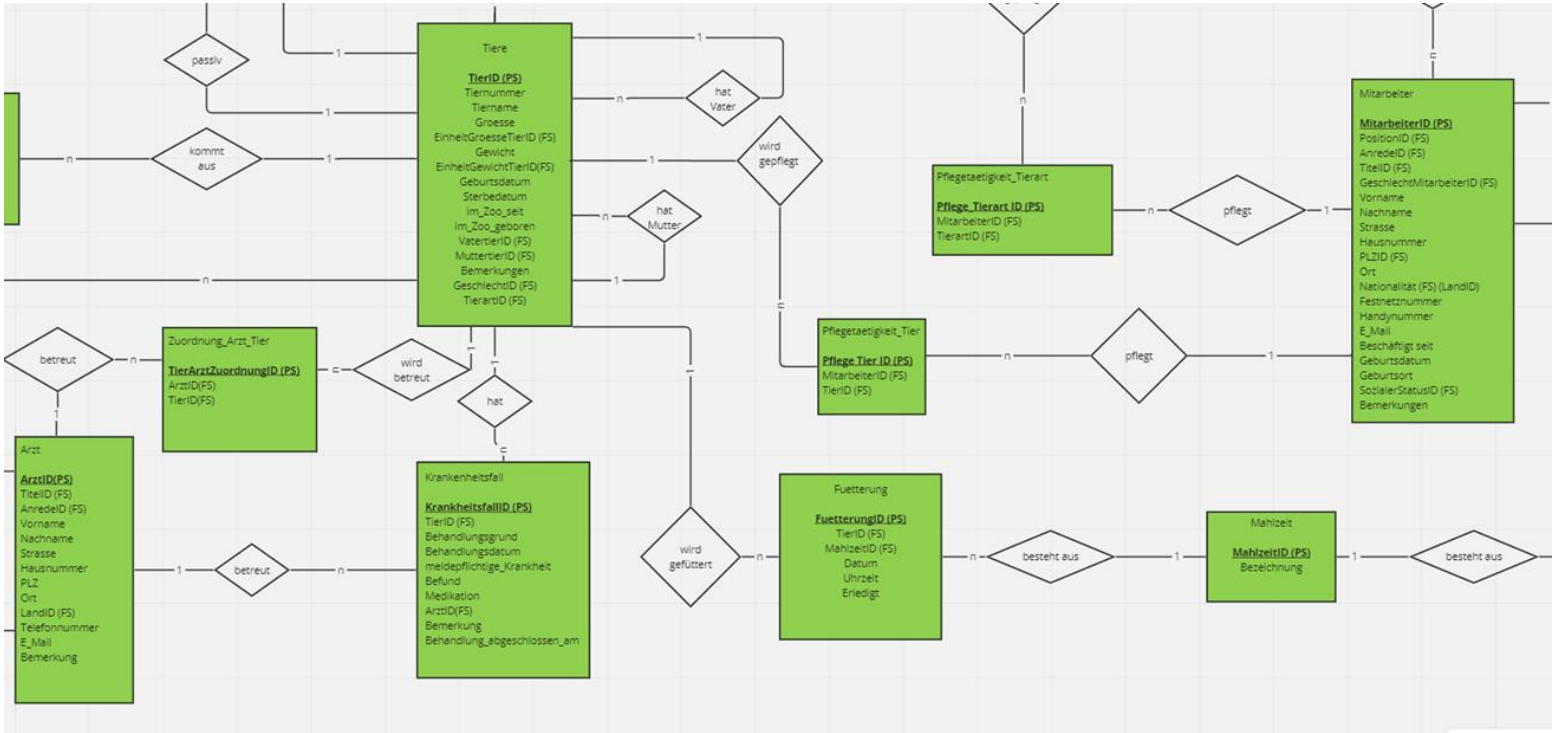
https://miro.com/app/board/uXjVM_y8W64=/

ERM: Referenztabellen

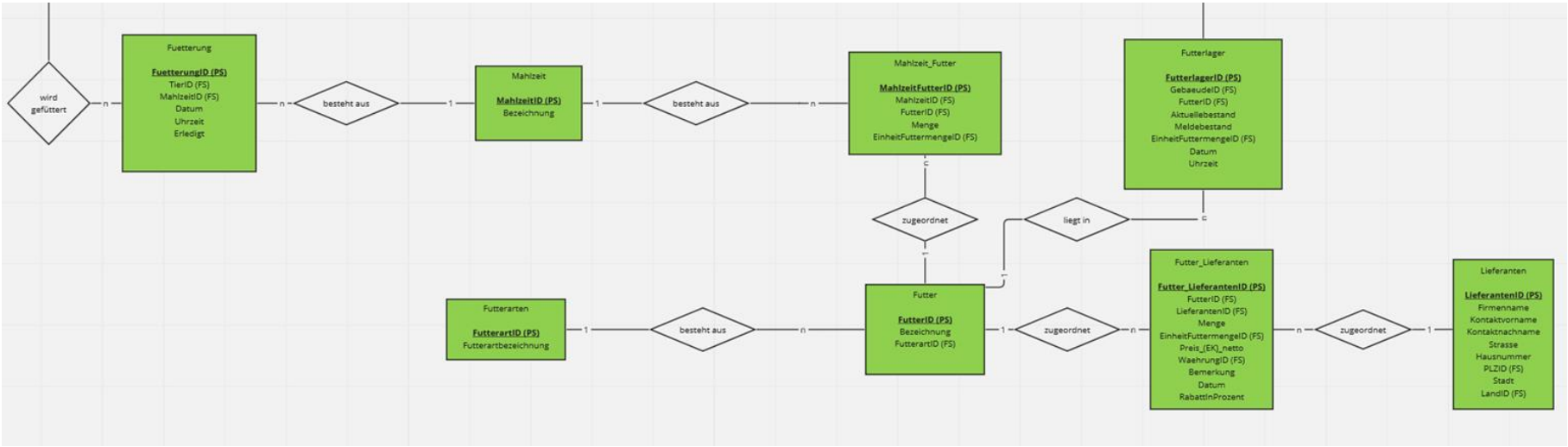
ERM: Tiere und Tierarten



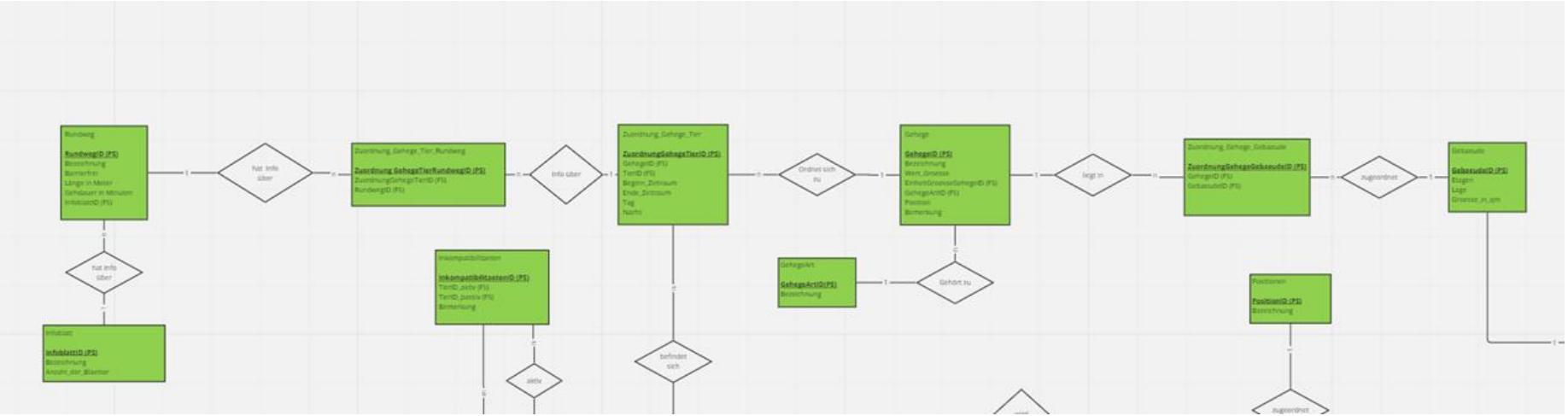
ERM: Ärzte und Mitarbeiter



ERM: Futter und Lieferanten



ERM: Gehege, Gebäude und Rundweg



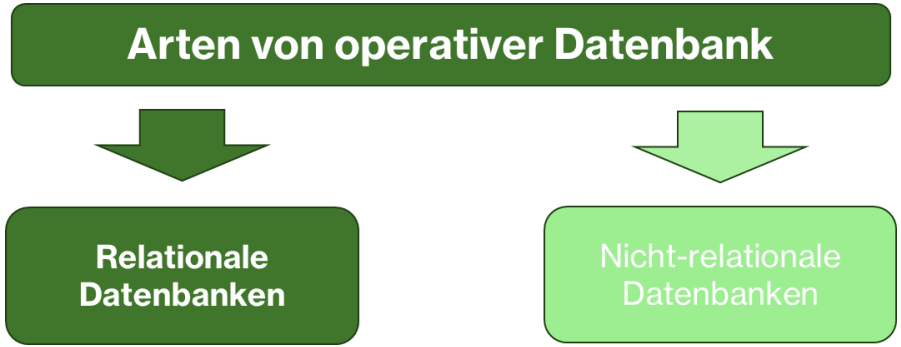
Operative Datenbank

Was ist eine operationelle (operative) Datenbank?

Eine operative Datenbank (ODB) ist eine elektronische Datenbank, die zur Speicherung und Verwaltung von Daten verwendet wird, die zur Unterstützung des täglichen Geschäftsbetriebs eines Unternehmens benötigt werden. Dieser Datenbanktyp speichert aktuelle Informationen über den aktuellen Stand des Unternehmens und wird daher verwendet, um Entscheidungen in Echtzeit zu treffen.

Operative Datenbank

Was ist eine operationelle (operative) Datenbank?



Relationale Datenbanken sind strukturiert und in Form von Tabellen organisiert, während nicht-relationale Datenbanken flexibler und nicht in einer bestimmten Struktur organisiert sind.

Operative Datenbank

Minimale Anforderungen an operativer Datenbank

- Struktur der Daten: Tabellen und Beziehungen sollten echte Geschäftsbeziehungen reproduzieren
- Integritätssicherung: Daten werden auf Korrektheit (bereits während der Eingabe) überprüft und Fehlmanipulationen verhindert
- Redundanzarmut: Normalisierung der Daten
- Datensicherheit: ungewollter Datenverlust wird durch interne Backup- und Prüfmechanismen verhindert
- Datenschutz: Zugriffskontrolle und spezifische Sichten sorgen für einen Zugang gemäß der Rechte des Nutzers

Operative Datenbank

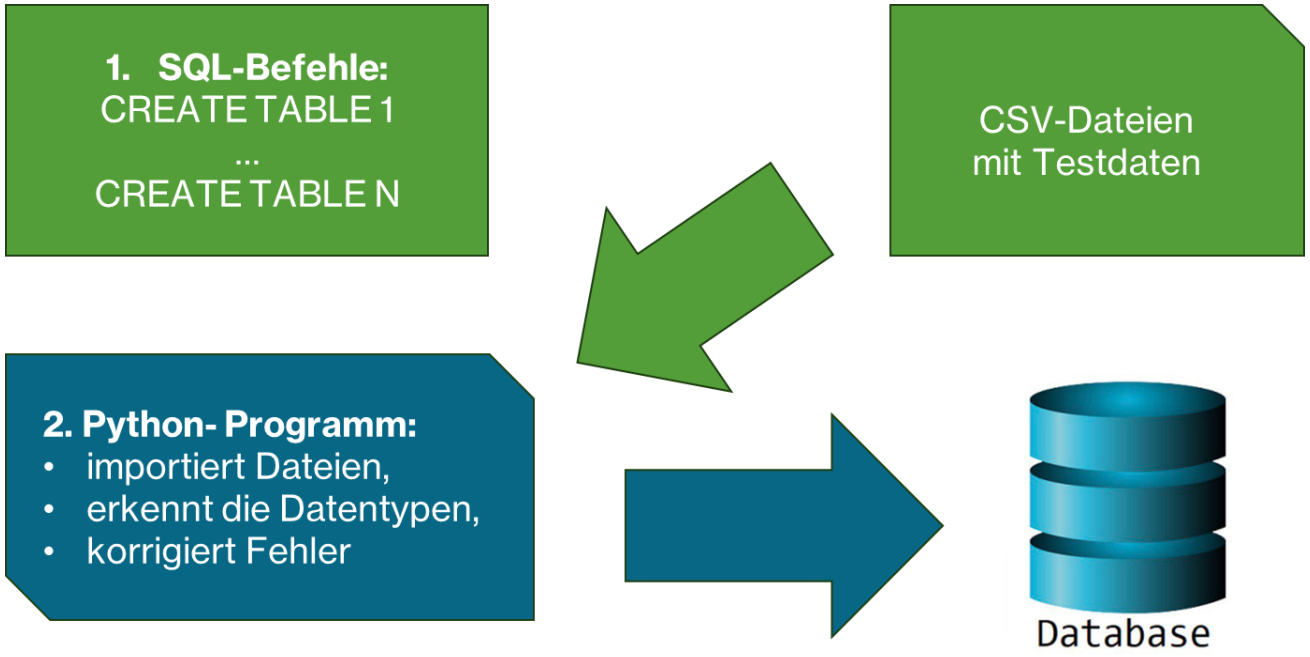
Auswählen eines Softtools. Warum SQLite?

SQLite ist einfache und optimale Lösung für den Einstieg:

- Es bietet viele der Funktionen von Standard-SQL und ist gleichzeitig kostenlos
- Es braucht wenig Platz für das Programm und Datenbanken
- Jeder, der sich mit SQL kennt, kann einfach verschiedene Abfragen stellen
- Mit mehreren Programmierungssprachen kann man Abfragen zu einer Datenbank automatisiert
- Später kann die Datenbank problemlos in eine andere SQL - Umgebung migriert werden

Operative Datenbank

Ausfüllung der Datenbank mit Testdaten



Data Dictionary

- Data-Dictionary dient der physischen Datenmodellierung [Wikipedia]
- Genaue Angaben zu
  - Tabellen und Datenfeldern
  - Primär- und Fremdschlüsselbeziehungen
  - Integritätsbedingungen
- DD wird auch Metadatenbank genannt [https://wikis.gm.fh-koeln.de/Datenbanken/Data-Dictionary]
  - „Metadaten oder Metainformationen sind strukturierte Daten, die Informationen über Merkmale anderer Daten enthalten. Auch Angaben von Eigenschaften eines einzelnen Objektes (beispielsweise „Personenname“).“ [Wikipedia]

Data Dictionary

- Entstehungsprozess nach dem Wasserfallmodell
  - Änderung im ERM → Datenbank → Data Dictionary
  - Auffälligkeiten bei Erstellung des Data Dictionary → DB → ERM
- Passives Data Dictionary [Wikipedia]
  - Manuelles Nachpflegen der Änderungen (unser Data Dictionary)
  - Aktualität nicht automatisch gewährleistet
- Zur Abgrenzung: aktives Data Dictionary [Wikipedia]
  - „Ein aktives Data-Dictionary reflektiert jederzeit den aktuellen detaillierten Stand des Datenmodells."
  - „Änderungen an der Struktur einer Datenbank können direkt in der Pflegeoberfläche des Data-Dictionary erfolgen, oder mit anderen Mitteln, zum Beispiel einem Kommandointerpreter einer DDL."
  - Aktualität automatisch gewährleistet

Data Dictionary

- Schwierigkeiten bei der Erstellung des Data Dictionary
  - Nachziehen der Änderungen in ERM und DB
  - Manche Dinge, die nicht passen, fallen erst im Data Dictionary auf
    - In der Regel sind das kleinere Sachen in der DB (aber immer Abgleich mit ERM)
    - Müssen dann wieder in DB und ERM eingepflegt werden
    - z.B. PLZ (Nummer) hat in der entsprechenden ERM-Tabelle gefehlt
- Aktuell halten des Data Dictionary lief sehr gut, da Kommunikation mit ERM-Maintainern (alle) und DB sehr gut und regelmäßig war

Data Dictionary

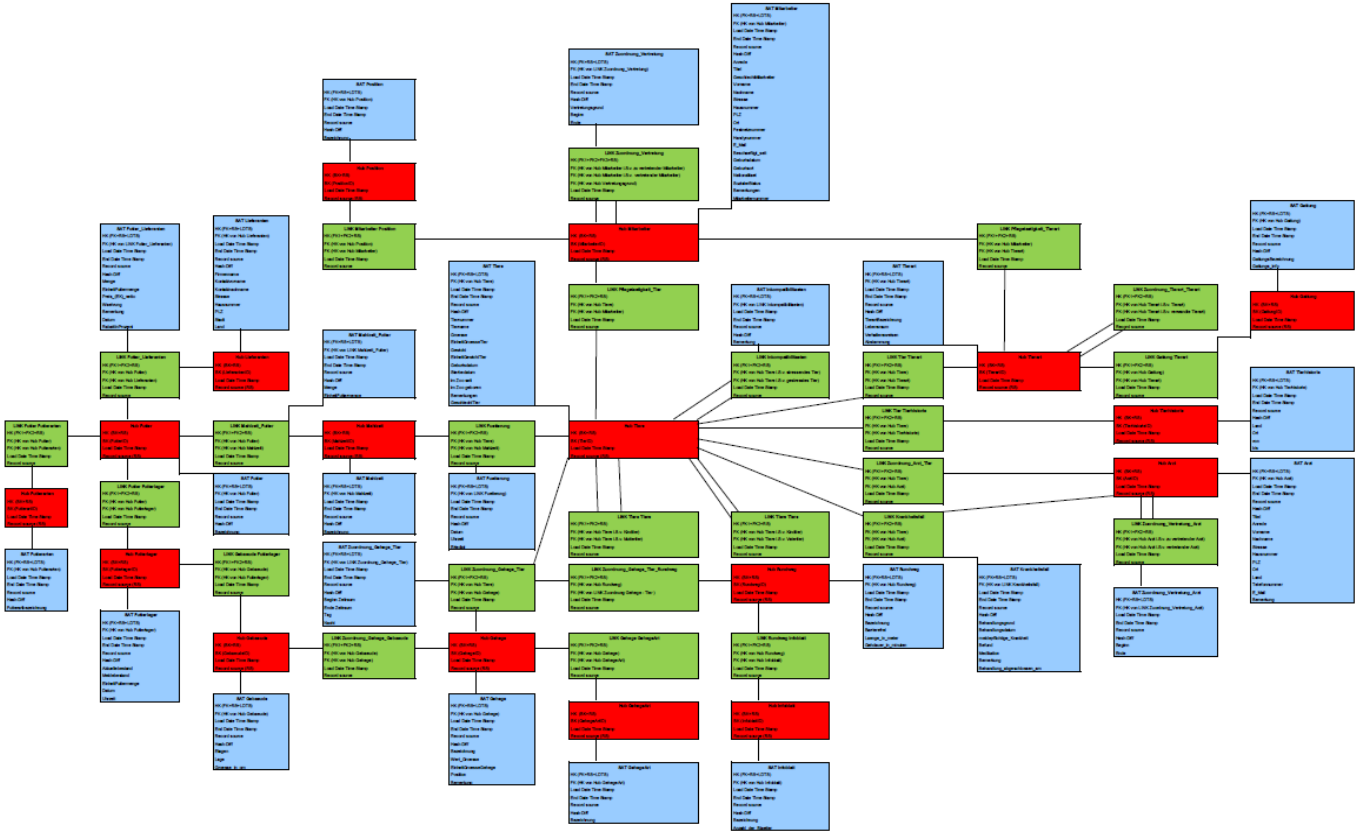
- 217 Spalten, 46 Tabellen
- [Zeige Data Dictionary]

Datawarehouse: Data Vault

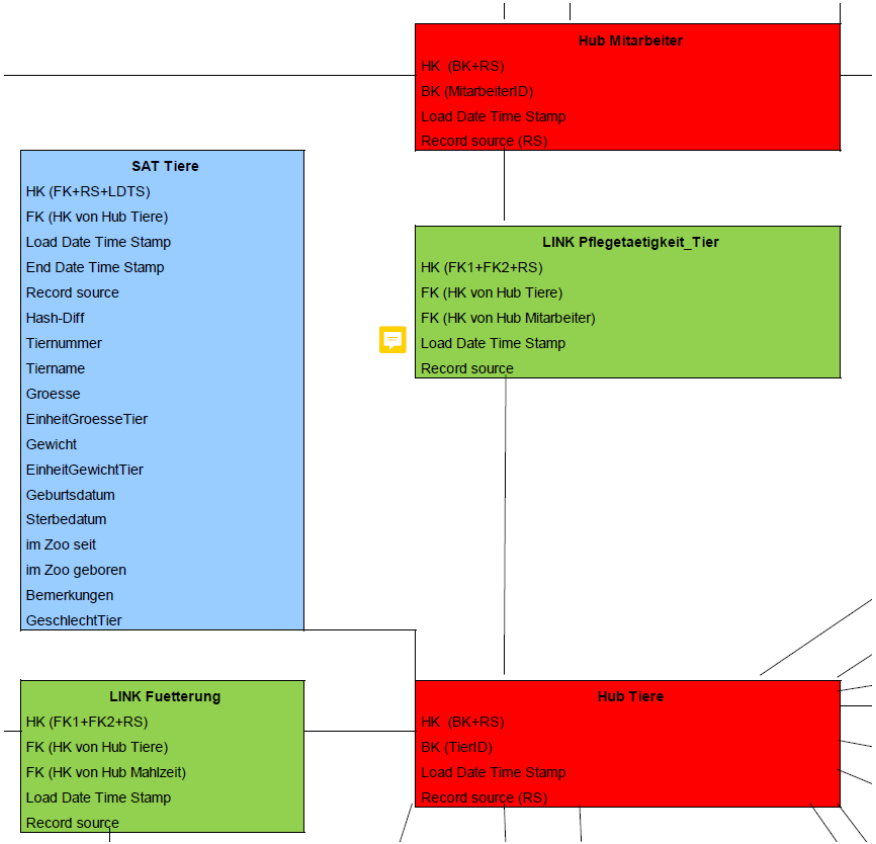
Warum Data Vault?

- Passend für wachsendes Unternehmen in den Anfängen der Modernisierung
  - Hohe Flexibilität
  - Skalierbar
- Einfache Struktur
- Nachteile kommen voraussichtlich nicht zum Tragen
- Alternative Galaxy Schema zu behäbig
  - Komplexe Struktur
  - Zeitintensive Erstellung
  - Positiv: Abfrageperformance

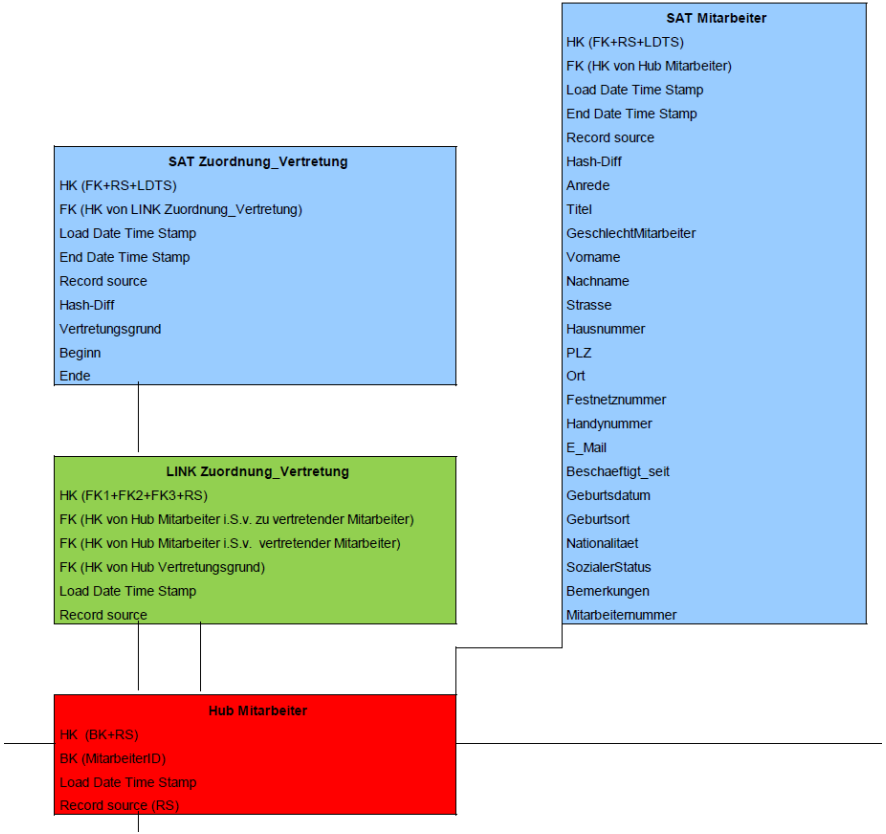
DATAWAREHOUSE: DATA VAULT



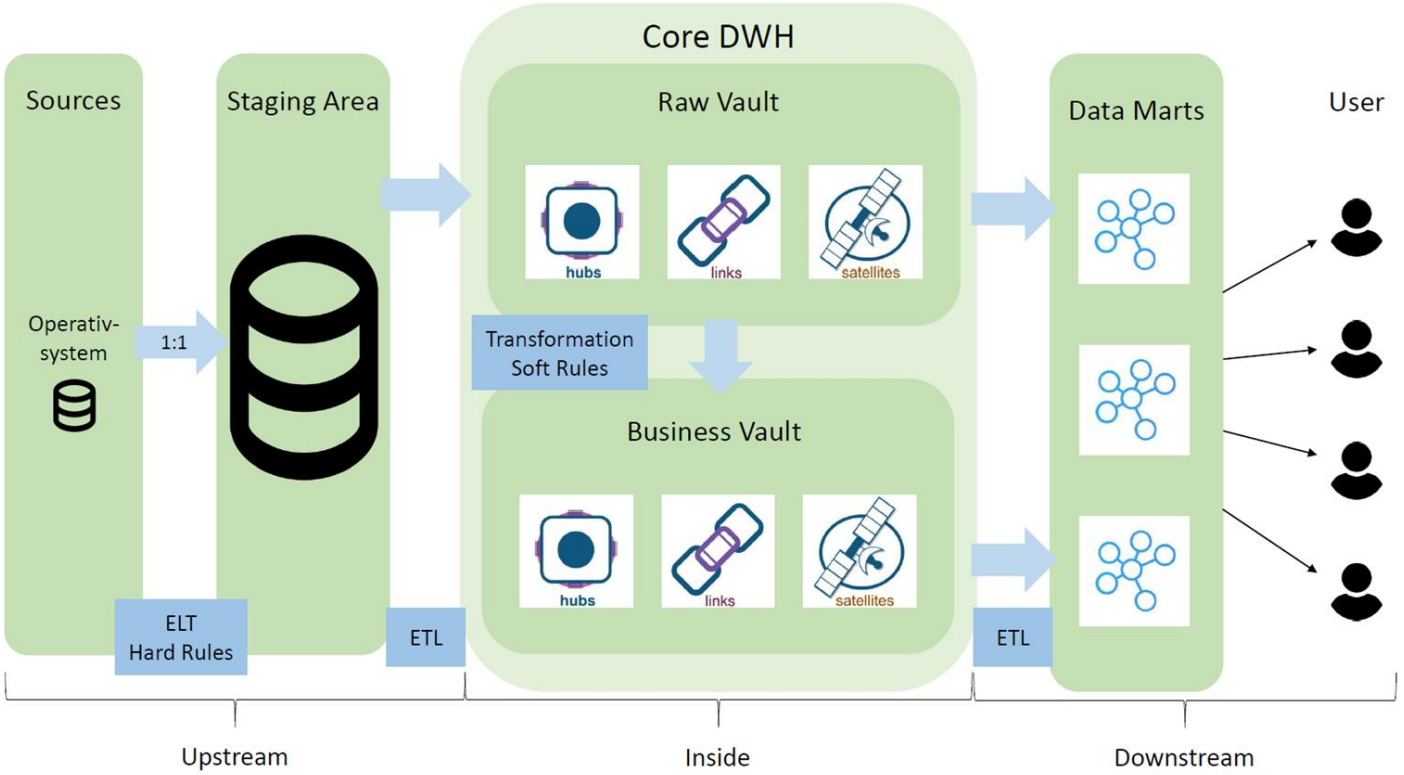
DATAWAREHOUSE: DATA VAULT



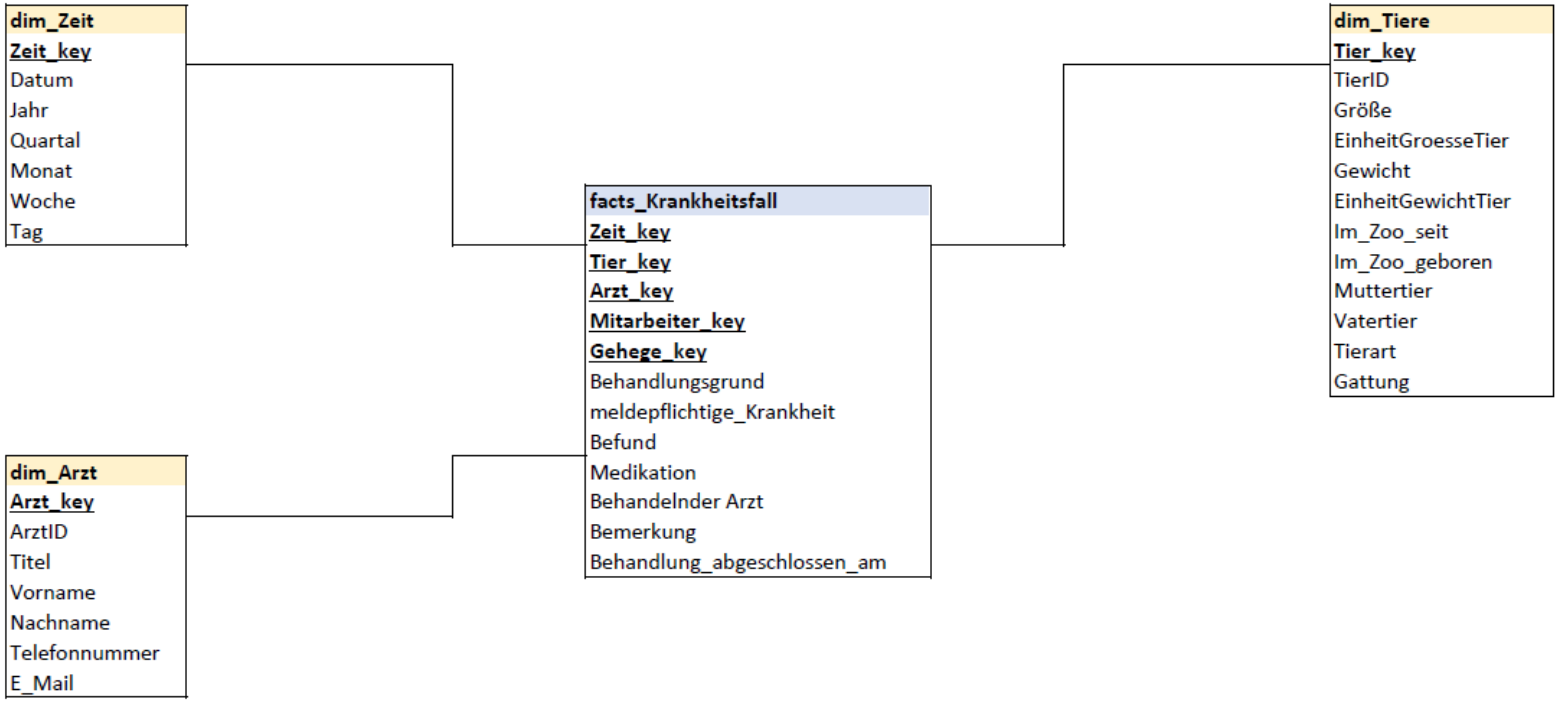
DATAWAREHOUSE: DATA VAULT



Datawarehouse: IT-Architektur



Data Mart: Krankheitsverlauf



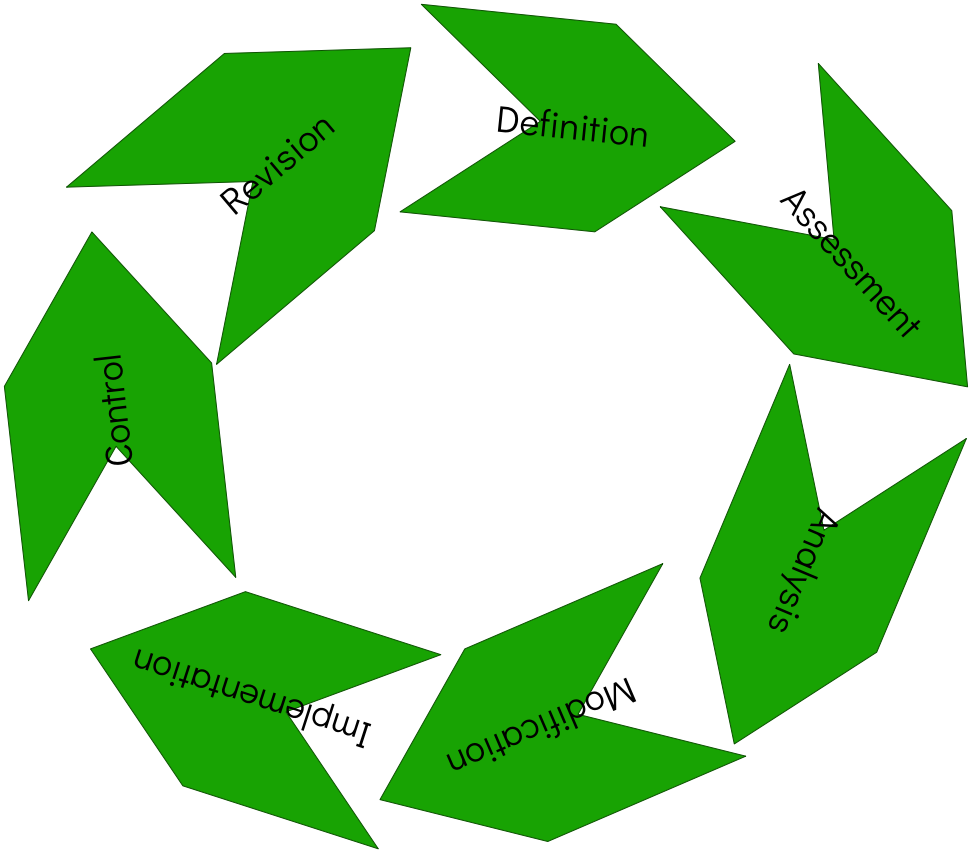
Verwendete Methoden, SW, Tools

- Miro für das ERM
- SQLite für die Datenbank
- Excel für das Data Dictionary
- Excel für das Datenmodell DWH
- PowerPoint für die IT Architektur DWH
- Word für das Dataquality Konzept

Data Quality Dimensions

- Assessment of old data
- Integration of legacy data
- Creating new data
- Combining the two in one system
- The role of the data vault
- The human in the loop
- Used cases
- Conclusion and outlook

DataQuality and Data Assessment



Requirements for data quality

- Data assessment, modification, standardization
- Digitalization and migration of data
- High data quality assurance (over 97%) at any time
- Definition and description of dimensions in DWS
- Listing of attributes
- Knowledge transfer and training

Data Quality Dimensions

- Assessment of old data
- Data modification and migration
- Creating new data
- Combining the two in one system
- The role of the data vault
- The human in the loop
- Used cases
- Conclusion and outlook

Assessment of old data

- Data profiling
- Data assessment
- Data migration
- Data modification and improving data quality
- Data monitoring

Data Assessment example

Animal

Attributes: animal ID, animal name, size, unit (size), weight, unit (weigh), date of birth, date of death, in zoo since, born in zoo, father ID, mother ID, gender ID, species ID, remarks

Some animal data is for some animals acquired in the past is missing the zoo documents. The data for missing units have been added. The remark section has been edited to specify what exactly was measure for the size. Subsequently the information of leaflets has been updated.

Animal assignments to at least one zookeeper and one veterinarian was checked. In cooperation with the zoo administration the assignment has been completed, wherever it was blank. The incompatibility of animals has been updated based on the reports of zookeepers and updated in the animal files. In a compound of n animals there is a maximum number of C(n,2) possible incompatibilities to be checked. We wrote a note in compounds in which some compatibility relations have not been investigated yet. This information can be added once this investigations are finalized by the corresponding responsible zoologist.

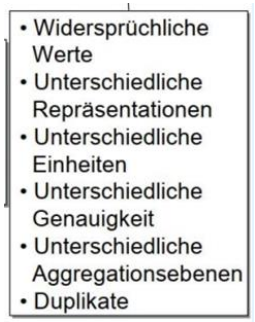
Data assessment results

Data assessment results

	Initial state	Current state	Remarks
Completeness	average	good	Some data should be measured and check before completion
Correctness	good	very good	
Timelineness	average	good	The aggregation of data must be tracked more regularly and consequently
Precision	average	very good	
Redundancy	average	very good	
Relevance	average	very good	
Uniformity	average	very good	
Consistency	good	very good	
Comprehensibility	average	very good	

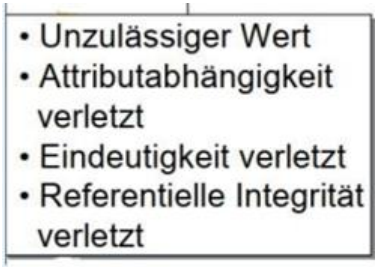
Creating new data

- Drop down menus in GUI
- Cross-checking for e.g. zip code and city or animal and animal family
- Checking for duplicates when entering employees
- Checking care times for overlap



Combining the two in one system

- Luckily not that big of a problem with homogenous data sources → e.g. not a kg lbs problem
- Careful planning with naming and structuring → snake vs camel case, acronyms etc.
- Detail level needs to be the same
- Data dictionary important tool here



The role of the data vault

- Data vault is detail-oriented, history tracing and allows to link unique tables
- Flexible, scalable, consistent, and adjustable system
- Cheap to maintain → money, time and energy can go into keeping high data quality level
- Expansion of system is simple, no adjustements needed → ensures data consistency and therewith quality in the future
- Speed advantage due to parallel loading possibilities not yet relevant, however might become important with ever growing system

The human in the loop

- Access decisions are important
- Guiding people with Gui, requirements, triggers and cross-checking is good
- Same goes for automization
- However for the points a person will have to interact with the system training is key

USE CASES

Used case	Neues Tier anlegen / in diesem Zuge ggf. auch neue Gattung / Tierart anlegen
Data and scheme	For animal family the right level needs to be checked before entering a new one e.g. mammal is not the same level of detail as rodent or marsupial For a new animal it is important to insert the entire history of it if it was not born in the Primasens zoo (we will insert a trigger for this case that reminds the user if a new animal is inserted and the born in zoo option is not checked) A newly inserted animal family and/or a novel animal will both remind the user to assign a person that is responsible for this new item in the data base The system will automatically send out this information to assigned caretaker via email to let them know of their new task and giving the option to schedule caretaking activities right away The system will also require incompatibility info, any medical history data, feeding facts and enclosure information at the spot and remind the user to insert it
Human in the loop	Only the back office and caretakers will have access to this option
Additional remarks	

DATA QUALITY

Conclusion and outlook

- Assessment of printed zoo data show a potential for improvement
- The initial data quality has been improved
- The improved data have been digitalized and migrated to a DWS
- The high quality of the new data system can be maintained by using a data vault
- A series of user oriented training courses are offered in order to transfer the knowledge to employees

Nächste Schritte

Operativ

- Daten für Datenbank auswählen (juristische Fragen klären)
- Front-End-Lösung erstellen, die die Anzahl der Eingabefehler minimiert
- Altdaten hochladen oder über Front-End-Lösung eingeben (analoge Quellen)
- Eingabe aller neuen geschäftlichen Daten erfolgt ausschließlich in die operative Datenbank

Strategisch

Migration von Daten aus:

- Buchung Onlinetickets
- Kampagnen und Werbung
- Verkauf aus Gastro und Merchandise
- Besucherdaten

Danke für eure Aufmerksamkeit

Fragen?

Zuordnung der Arbeitspakete

- Datenmodell (ERM) – alle
- Datenbank Operativsystem – Vsevolod Dorskiy
- Datenbankdokumentation / Data Dictionary – Jerg Jaisle
- Datenmodell Datawarehouse und Datawarehouse IT Architektur – Tobias Gründer
- Dataquality Konzept – Tobias Berger und Mohammad Hessan
- Präsentation - alle