

# Der Anforderungs- und Abnahmeprozess in einer modellorientierten DWH-Entwicklung

White Paper der solvistas GmbH

Thomas Neuböck  
Jürgen Raab  
Melissa Schmidt

Linz, April 2022

## 1 Einleitung

In der Entwicklung von Business Intelligence (BI) und Data Warehouse (DWH) Systemen stellt die Konfiguration eines geeigneten Prozesses zum Requirements Engineering eine besondere Herausforderung dar. Anforderungen werden kooperativ, iterativ und explorativ mit dem Fachbereich entwickelt. Die Systementwicklung kann in einem agilen Stil durchgeführt werden. Die Grenzen zwischen Anforderungsanalyse, Datenanalyse (Data Profiling, Data Assessment) und Systementwicklung verschwimmen. Eine Projektplanung mit einem längeren Zeithorizont ist vor allem in Unternehmen, die klassische Vorgehensweisen im Sinne eines Wasserfallmodells gewohnt sind, wird nahezu unmöglich und erzeugt Unverständnis im Management, insbesondere wenn es um Abnahme von (vorher eigentlich gar nicht klar definierbare) Lieferobjekten geht. Aussagen wie „Dieses Tortendiagramm kostet 1 Mio. Euro!“ zeigen das Dilemma.

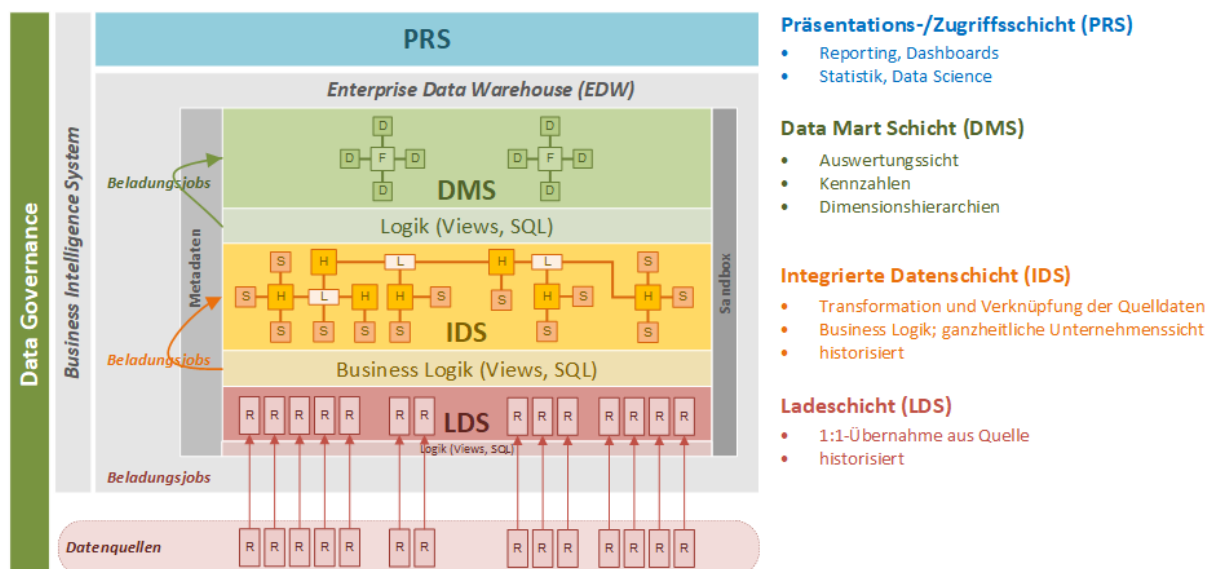
In diesem White Paper wird ein Vorschlag für einen möglichen Abnahmeprozess in der DWH-Entwicklung vorgestellt. Da es vor allem bei einem größeren fachlichen Thema oft schwierig ist, die zugrundeliegenden Datenquellen für die Umsetzbarkeit von Anforderungen richtig einzuschätzen (und darauf eine seriöse Planung aufzusetzen), wird vorgeschlagen, den Abnahmeprozess in überschaubare Teile zu gliedern.

Dieser Vorschlag geht in die Richtung, dass eine Gliederung in mehrere Abnahmen sich einerseits an der DWH-Architektur und andererseits an einer iterativen und inkrementellen Vorgehensweise in der Umsetzung orientiert. Die zugrundeliegende modellorientierte Vorgehensweise (drei Datenschichten der Architektur und drei Modellebenen) stellt dabei eine wichtige Basis dar. Darauf wird der Vollständigkeit halber (als Hintergrundinformation) in den beiden folgenden Unterabschnitten noch genauer eingegangen. Diese Gliederung ermöglicht zudem eine feinere und zielführendere Projektsteuerung.

## 1.1 Architektur

Das hier betrachtete DWH besteht aus **drei Datenschichten**, wobei die Reihenfolge der nachfolgenden Auflistung bewusst in der Reihenfolge eines Abnahmeprozesses gewählt wurde:

- Ladeschicht (LDS)
- Integrierte Datenschicht (IDS)
- Data Mart Schicht (DMS)



Die **Ladeschicht (LDS)** entspricht einer Persistent Staging Area, in der die Daten aus den Quellen annähernd 1:1 übernommen und historisiert (mit Ladezeitstempel) abgelegt werden. Die Datenübernahme folgt dabei einer Delta-Logik, d.h. es werden nur neue oder veränderte Datensätze übernommen und mit einem technischen Ladezeitstempel versehen. Diese Deltaermittlung erfolgt entweder über geeignete Zeitattribute der Datenquelle (z.B. technisches Erstellungs- und Änderungsdatum), falls diese dort verfügbar sind, oder über einen Vollabgleich (Vergleich des Datenbestands in der Quelle mit dem im DWH).

In der **integrierten Datenschicht (IDS)**, welche einem Core Data Warehouse entspricht, werden die Quelldaten aus der LDS übernommen und in Anlehnung an Data Vault 2.0 strukturiert (Dan Linstedt, Bill Inmon). Daten der LDS werden in der IDS historisiert als Raw Vault aufbereitet und falls sinnvoll mit zusätzlicher Logik (Business Vault) angereichert. Die IDS ist flexibel hinsichtlich Erweiterungen und bietet die potenzielle Möglichkeit Daten aus verschiedenen Quellen zu verknüpfen (neben der Historisierung zwei wichtige Grundanforderungen an ein zeitgemäßes Enterprise Data Warehouse).

Die **Data Mart Schicht (DMS)** stellt Daten der IDS einem Business Intelligence (BI) Tool für Auswertungen und Analysen durch Anwender\*innen zur Verfügung. Die Daten der DMS werden entweder im einfachsten Fall als flache Tabellen zur Verfügung gestellt oder im Regelfall als dimensionales Modell im Star Schema (nach Ralph Kimball) aufbereitet.

Der unten definierte Abnahmeprozess bezieht sich diese drei Datenschichten. Die Präsentations- und Zugriffsschicht (PRS) umfasst die Umgebung des BI-Tools und wird in diesem Abnahmeprozess nicht betrachtet.

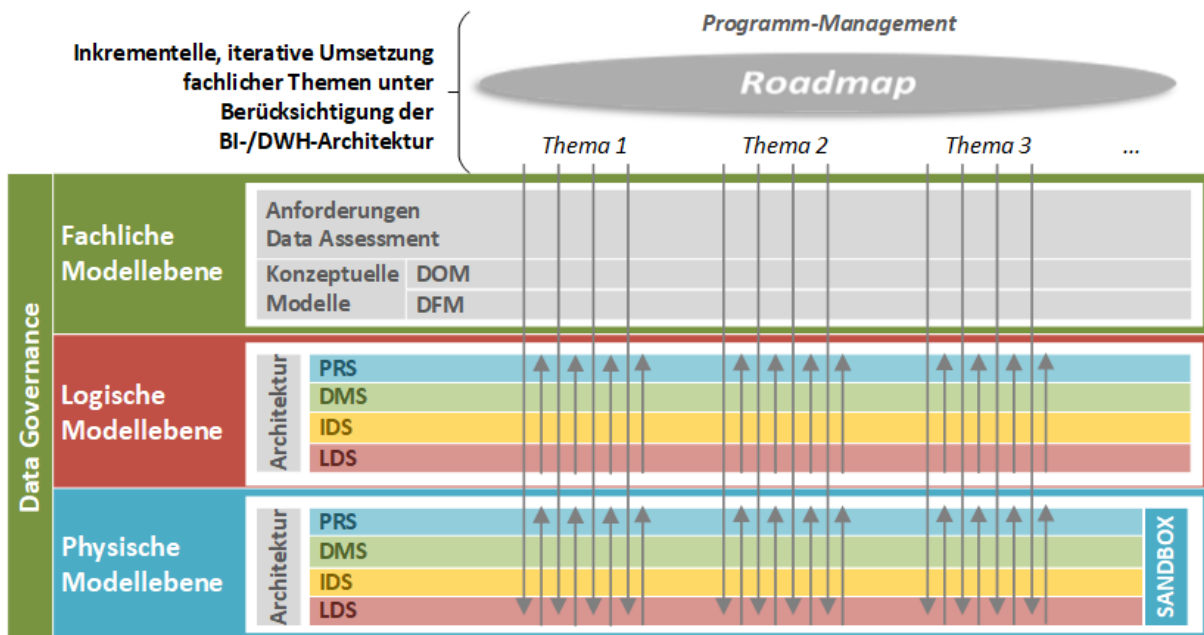
Als logische Abnahmezeitpunkte bieten sich Übergänge in die einzelnen Datenschichten an:

- Beurteilung der Datenquelle hinsichtlich Verwendbarkeit zur Umsetzung der fachlichen Anforderungen (Abnahme des Data Assessment der Datenquelle)
- Abnahme übernommener Objekte in der LDS
- Beurteilung der übernommenen Objekte in der LDS hinsichtlich Verwendbarkeit zur Umsetzung der fachlichen Anforderungen (Abnahme des Data Assessment der Objekte in der LDS)
- Abnahme übernommener Raw-Vault-Objekte in der IDS
- Beurteilung der übernommenen Raw-Vault-Objekte in der IDS hinsichtlich Verwendbarkeit zur Umsetzung der fachlichen Anforderungen (Abnahme des Data Assessment der Raw-Vault-Objekte in der IDS hinsichtlich Spezifikation der Business-Logik in der IDS und hinsichtlich Spezifikation der DMS)
- Abnahme der Business-Vault-Objekte der IDS
- Abnahme der Objekte der DMS

Eine genauere Beschreibung dazu findet man im Abschnitt 2 bzw. werden diese Überlegungen auch im folgendem Unterabschnitt 1.2 im Sinne einer iterativen und inkrementellen Vorgehensweise herangezogen.

## 1.2 Vorgehensweise

Die DWH-Umsetzung umfangreicher Fachthemen erfolgt in der Regel iterativ und inkrementell, was in der folgenden Abbildung ausgedrückt wird:



Fachliche Anforderungen werden aufgenommen, Datenquellen werden identifiziert, analysiert und hinsichtlich Verwendbarkeit zur Umsetzung der fachlichen Anforderungen beurteilt (Data Assessment der Datenquelle). Quelldaten werden schrittweise in die LDS und in die IDS (Raw Vault) übernommen. Nach der Übernahme in die LDS bzw. als Raw Vault in die IDS erfolgt jeweils eine weitere Beurteilung der Daten hinsichtlich Umsetzung der

fachlichen Anforderungen (Data Assessment der LDS-Daten<sup>1</sup>, Data Assessment der Raw-Vault-Daten<sup>2</sup>).

Anforderungen eines umfangreicheren Fachthemas können in mehrere Teilbereiche gegliedert werden. Häufig können die Anforderungen zu Beginn eines Vorhabens kaum vollständig identifiziert und hinsichtlich Umsetzung bewertet werden. Die Teilbereiche eines Fachthemas werden in einem iterativen Prozess inkrementell umgesetzt. Nicht selten stehen erst nach Fertigstellung eines Teilbereiches genügend Kenntnisse zur Verfügung, um weitere Teilbereiche mit ausreichender Genauigkeit zu planen und spezifizieren zu können. Die Gliederung eines umfangreichen Fachthemas in verschiedene Teilbereiche (Umsetzungsphasen) erfolgt nach fachlichen Gesichtspunkten und bilden die Voraussetzung zur Planung einer übergeordneten Roadmap.

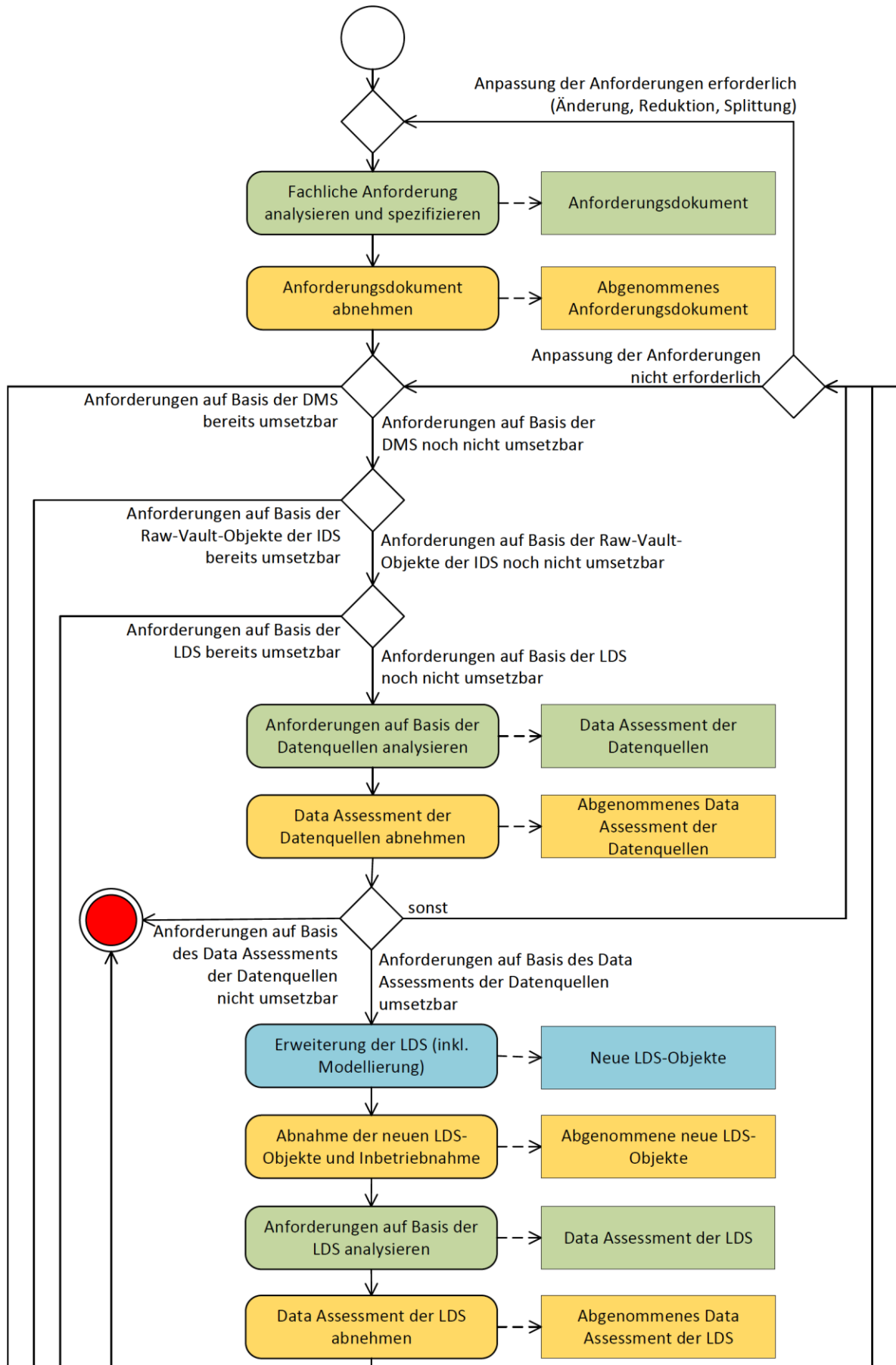
## 2 Abnahmeprozess

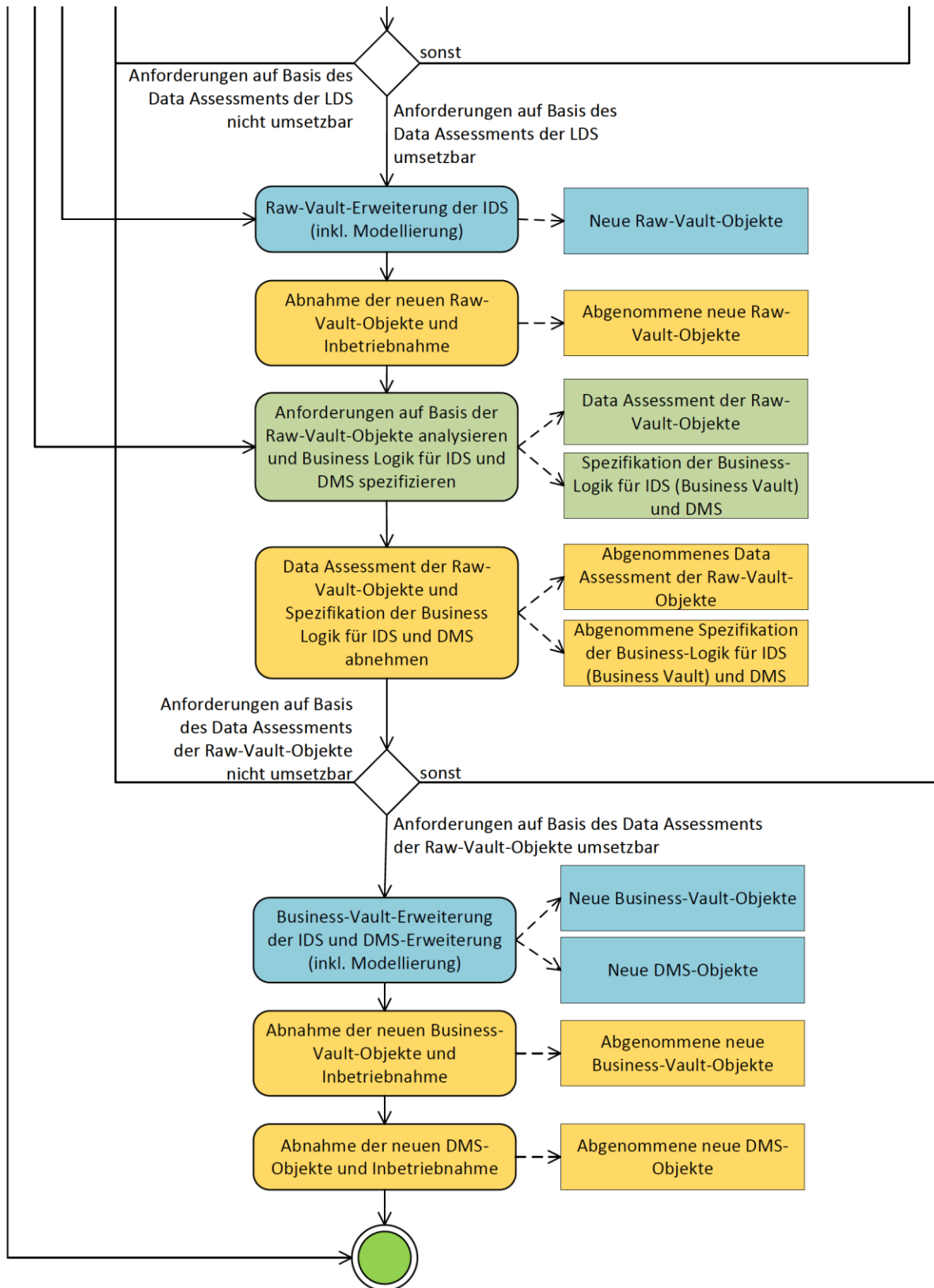
In diesem Abschnitt wird ein Abnahmeprozess vorgeschlagen, der nicht eine einzelne Gesamtabnahme am Ende eines Gesamtvorhabens vorsieht, sondern, indem einzelne Teile in der gesamten Entwicklung beurteilt und abgenommen werden. Dies bietet darüber hinaus die Möglichkeit, die Umsetzung eines umfangreicheren Fachthemas feingranularer zu steuern. Das folgende Ablaufdiagramm stellt diesen Abnahmeprozess im Detail dar:

---

<sup>1</sup> In der LDS können nach einem gewissen Zeitraum die Daten auch hinsichtlich Delta-Beladungen überprüft werden.

<sup>2</sup> Auf Basis von Raw-Vault-Objekten in der IDS können Daten historisch zu bestimmten Stichtagen (abhängig von den fachlichen Anforderungen) analysiert werden.





Im ersten Schritt werden fachliche Anforderungen analysiert, spezifiziert und in einem Anforderungsdokument festgehalten. Dieses Dokument bildet die Basis und ist als gemeinsames Commitment zwischen Kunde (Fachbereich) und Umsetzung zu verstehen und vom Kunden abzunehmen (**abgenommenes Anforderungsdokument**).

Erst danach werden weitere Schritte in Richtung Umsetzung eingeleitet. Sind dabei die fachlichen Anforderungen bereits auf Basis der bestehenden DMS umsetzbar, sind innerhalb des PVDWH's keine weiteren Umsetzungsschritte erforderlich.<sup>3</sup> Im anderen Fall ist zu überprüfen, ob die Anforderungen auf Basis der Raw-Vault-Objekte in der IDS implementiert werden können. Ist dies der Fall, können die Anforderungen auf Basis der Raw-Vault-Objekte analysiert und die Business Logik für die IDS und die DMS spezifiziert werden. Ansonsten wird überprüft, ob die fachlichen Anforderungen zumindest auf Basis der LDS umgesetzt werden können. Trifft dies zu, folgt die Erweiterung der IDS um die erforderlichen Raw-Vault-Objekte. Im anderen Fall ist eine Umsetzung in allen Datenschichten erforderlich. Diese wird in der weiteren Beschreibung genauer ausgeführt.

Zunächst sind die fachlichen Anforderungen auf Basis der verfügbaren Informationen über die relevanten Datenquellen genauer zu analysieren. Die Verwendbarkeit der Datenquellen zur Umsetzung der Anforderungen ist zu beurteilen und entsprechend zu dokumentieren (Data Assessment der Datenquellen). Diese Dokumentation ist in Folge dessen wiederum einer Abnahme zu unterziehen (**abgenommenes Data Assessment der Datenquellen**). Diese Abnahme stellt somit ein weiteres Commitment zwischen Kunde und Umsetzung dar. Darin enthalten ist auch eine vom Kunden getroffene Entscheidung über die weitere Vorgehensweise, wobei drei grundsätzliche Kundenentscheidungen möglich sind: (1) Auf Basis der Analyse der Datenquelle erscheint die Umsetzung der Anforderungen möglich zu sein und es wird vom Kunden entschieden, mit der Umsetzung der LDS fortzusetzen. (2) Auf Basis der Analyse der Datenquelle erscheint die Umsetzung der Anforderungen nicht (sinnvoll bzw. zielführend) möglich zu sein und es wird vom Kunden entschieden, die Umsetzung abzubrechen. (3) Auf Basis der Analyse der Datenquelle ist es sinnvoll bzw. zielführend die Anforderungen neu zu bewerten oder abzuändern und es wird vom Kunden entschieden, entweder die zur nochmaligen Bewertung der bereits vorhandenen Daten in den Datenschichten (DMS, IDS, LDS) zurückzukehren oder die Anforderungen nochmals zu analysieren und abzuändern (d.h. Anforderungen werden abgeändert oder reduziert; oder es wird eine Anforderung in mehrere Anforderungen aufgeteilt und ggf. werden daraus entstehende Anforderungen in einer späteren Projektphase betrachtet und aus dem Scope der aktuellen Projektphase entfernt).

Wurde seitens des Kunden im abgenommenen Data Assessment der Datenquelle festgehalten, dass die Anforderungen auf Basis der Analyse der Datenquellen umsetzbar sind, wird mit der Erweiterung der LDS (Modellierung und Implementierung) fortgesetzt. Es entstehen neue LDS-Objekte, die vom Kunden abgenommen und auf Basis dessen in Betrieb genommen werden (**abgenommene neue LDS-Objekte**).

Auf Basis der abgenommenen und in Betrieb genommenen LDS-Objekte können die fachlichen Anforderungen hinsichtlich Umsetzbarkeit einer weiteren genaueren Analyse unterzogen werden. Zum Beispiel können Anforderungen hinsichtlich mehrere Delta-Beladungen über einen passenden längeren Zeitraum hinweg genauer untersucht werden. Die daraus erhaltenen Erkenntnisse werden in einem Data Assessment der LDS festgehalten und dem Kunden zur Abnahme unterbreitet (**abgenommenes Data Assessment der LDS**). Diese Abnahme stellt wiederum ein weiteres Commitment zwischen Kunde und Umsetzung dar und enthält analog zum abgenommenen Data Assessment der Datenquellen eine vom Kunden getroffene Entscheidung über die weitere

---

<sup>3</sup> Anmerkung: Umsetzungen in der Zugriffs- und Präsentationsschicht werden in diesem Abnahmeprozess nicht betrachtet.



Vorgehensweise. Die vorgegebenen drei Entscheidungsmöglichkeiten sind analog zu denen auf Basis des abgenommenen Data Assessments der Datenquellen zu betrachten.

Entscheidet sich der Kunde für eine Fortsetzung der Umsetzung, folgen die erforderlichen Raw-Vault-Erweiterungen in der IDS. Die daraus entstehenden Raw-Vault-Objekte sind wiederum vom Kunden abzunehmen und werden auf Basis dessen in den Betrieb übergeführt (**abgenommene neue Raw-Vault-Objekte der IDS**).

Auf Basis der abgenommenen und in Betrieb befindlichen Raw-Vault-Objekte in der IDS können die fachlichen Anforderungen hinsichtlich Umsetzbarkeit einer weiterführenden Analyse unterzogen werden. Beispielsweise ist es nun möglich, die Umsetzbarkeit von Anforderungen hinsichtlich stichtagsbezogenen Auswertungsmöglichkeiten genauer beurteilen zu können. Damit ist es nun auch möglich die Spezifikation der Business Logik für die IDS und die darauf aufbauende Spezifikation der DMS genau festlegen zu können. Daraus entstehen zwei Dokumentationen, die vom Kunden abzunehmen sind: **abgenommenes Data Assessment der Raw-Vault-Objekte der IDS** und **abgenommene Spezifikation der Business-Logik für die IDS (Business Vault) und DMS**. Das abgenommene Assessment der Raw-Vault-Objekte enthält analog zum abgenommenen Data Assessment der Datenquellen bzw. zum abgenommenen Data Assessment der LDS eine vom Kunden getroffene Entscheidung über die weitere Vorgehensweise. Die vorgegebenen drei Entscheidungsmöglichkeiten sind analog zu diesen zu betrachten.

Wird seitens des Kunden die Fortsetzung der Umsetzung entschieden, werden die Business-Vault-Objekte der IDS und die DMS-Objekte implementiert. Diesen werden wiederum vom Kunden abgenommen: **abgenommene neue Business-Vault-Objekte der IDS** und **abgenommene neue DMS-Objekte**. Auf Basis dieser Abnahmen werden diese Objekte in Betrieb genommen. Mit der Abnahme und der Inbetriebnahme der DMS-Objekte ist entweder ein gesamtes Fachthema oder ein fachlicher Teilbereich daraus vollständig umgesetzt, abgenommen und befindet sich vollständig im Produktivbetrieb des DWH's.