



Leistungsbeschreibung

Anlage B3

Software zur Datenbewirtschaftung
Stand: 11.09.2026

Inhalt

Vorwort.....	4
1 Ausgangslage: Datenbewirtschaftungsprozesse im WIdO	5
2 Kurzbeschreibung des Beschaffungsgegenstands und Ziel der Leistungserbringung	6
2.1 Anforderungen an die Software	7
2.2 Beispiele als Ergänzung zum Kriterienkatalog	10
2.2.1 Beispiel zu Deployen Teil 1 Muss-4.1.1:.....	10
2.2.2 Beispiel zu Protokollierung/Monitoring Teil 1 Muss-4.2.2	10
2.2.3 Beispiel zu Logging und Monitoring Teil 1 Muss-4.2.1	11
3 Nutzung der Software:	12
4 Installations- und Betriebsbedingungen für IT-Systeme im AOK- Bundesverband	13
4.1 Entwicklungsumgebung	13
4.2 Serverkomponente	13
4.3 Softwareanforderungen	14
4.4 Anforderungen an IT- und Datensicherheit	14
4.5 Mandantentrennung.....	14
5 Unterstützung bei der Inbetriebnahme	16
Abbildungsverzeichnis	17
Tabellenverzeichnis	18

Vorwort

Das Wissenschaftliche Institut der AOK (WIdO), als eine Geschäftsführungseinheit des AOK-Bundesverbandes, bietet fundierte Expertise über alle Leistungsbereiche der gesetzlichen Krankenversicherung und nimmt vor allem die Patientenversorgung in den Blick. Das WIdO versteht sich als Bindeglied zwischen Wissenschaft und Praxis. Das heißt einerseits, wissenschaftliche Erkenntnisse für die Praxis nutzbar zu machen, und andererseits, Fragen aus der Praxis auf die wissenschaftliche Agenda zu bringen. In beide Richtungen setzt das WIdO seit vielen Jahren wichtige Impulse. Damit das gelingt, werten die WIdO-Expertinnen und Experten jedes Jahr viele Millionen Datensätze zur Gesundheitsversorgung in Deutschland unter Qualitäts- und Wirtschaftlichkeitsaspekten aus.

Eine wesentliche Grundlage für die Aufgaben des WIdO sind empirische Daten auf Basis der über 28 Millionen AOK Versicherten. Basierend auf diesen Daten bietet das WIdO zahlreiche Produkte für das AOK-System sowie in der öffentlichen Nutzung an. Hierzu zählen Anwendungen und Webseiten aber auch gesetzlich verpflichtende Datenlieferungen für das AOK-System. Externe Kooperationen für Politik und Wissenschaft basieren ebenfalls auf diesen Datengrundlagen.

Das Data Warehouse des WIdO im Forschungsbereich „Integrierte Daten und Analysen“ beherbergt in der produktiven IBM Db2 Datenbank derzeit 72 TB komprimierte Daten. Das bisherig eingesetzte Tool wird nicht mehr gewartet, so dass ein neues Tool beschafft werden muss.

Die neue Softwarelösung soll als Standard für alle weiteren Forschungsbereiche des WIdOs sowie weitere Geschäftsbereiche im AOK-Bundesverband eingesetzt werden, so dass sich Anforderungen auch an Verarbeitungsprozesse anderer Datenquellen ergeben.

1 Ausgangslage: Datenbewirtschaftungsprozesse im WIdO

Das Data Warehouse im Forschungsbereich „Integrierte Daten und Analysen“ im WIdO besteht aus mehreren Datenbankschichten, um die Datenhaltung nachvollziehbar und die Daten gemäß den Datenschutzvorgaben vorzuhalten. Mit einer Schicht ist ein Datenbankschema gemeint, das themenbezogen zugeordnete Tabellen enthält. Die Datenmenge umfasst aktuell ca. 72 TB und wird regelmäßig aktualisiert und um weitere Daten erweitert.

Die Metadatenlogik für den ELT-Prozess ist in Steuerungstabellen der Datenbank hinterlegt. Über die Metadatentabellen und durch die, für die einzelnen Prozesse, abgelegten SQLs, bleibt der Gesamtprozess hoch transparent. Es werden vorwiegend Steuerungsflüsse erstellt und damit die volle Performance der Datenbank bei komplexen Datenverarbeitungsprozessen genutzt.

In den letzten 10 Jahren wurden zentrale Steuerungsflüsse aufgebaut, die ein schnelles Laden der zahlreichen Daten ermöglicht. Zentraler Gegenstand der ELT-Prozesse sind metadatengetriebene Kontrollflüsse. Das heißt, die Datenverarbeitung selbst findet auf der leistungsstarken Datenbank statt. Daneben gibt es für einige Datenkörper auch Datenflüsse, welche SQL-basierten Code erzeugen, um die im Kontrollfluss definierten Aufgaben zu erfüllen.

Zudem ist es möglich Data-Mining-Modelle mittels **graphischer** Data-Mining-Flows zu erstellen. Hierbei werden Daten aufbereitet, Data-Mining-Modelle erstellt und Visualisierungen von Datenmodellen erstellt, welche die Anwendung (oder Bewertung) der Modelle ermöglichen.

Das derzeitige Tool besteht aus

- Einer Entwicklungsumgebung mit einer Anbindung an ein Versionsverwaltungssystem
- Einer Serverkomponente für das Deployment der Flüsse inkl. einer kleinen Metadaten Datenbank.
- Einer entsprechenden browserbasierten Oberfläche zum Steuern bzw. Ausführen fertiger Prozesse.

In der Entwicklungsumgebung werden ELT-Prozesse von Entwicklerinnen oder Entwicklern des WIdO „gebaut“. Diese Prozesse werden dann in einer Serverkomponente deployed und stehen über eine Web-Oberfläche zur Verfügung. Diese kann auch von nicht Entwicklerinnen oder Entwicklern (**sogenannte Business User**) der Teams im WIdO genutzt werden, um Datenverarbeitungsprozesse zu starten.

2 Kurzbeschreibung des Beschaffungsgegenstands und Ziel der Leistungserbringung

Der Auftragnehmer hat dem Auftraggeber ein **Datenbewirtschaftungstool** zur Verfügung zu stellen. Das WIdO hat für verschiedene Projektkontexte vollautomatisierte hochkomplexe ELT-Prozesse etabliert und gepflegt. Es besteht die Notwendigkeit diese Prozesse nun möglichst schnell in eine neue Software zu überführen. Da die bisherigen Prozesse möglichst rasch in ein neues Tool überführt werden müssen, werden die Funktionalitäten des bisherigen Tools sowie die Anforderungen der bestehenden Prozesse aufgeführt. Diese sind auch durch die neue vom Auftragnehmer zu liefernde Softwarelösung abzudecken.

Zwischenzeitlich sind neue Anforderungen an die Datenverarbeitungsprozesse und an die **Qualitätssicherung** der Daten, sowie Data-Mining Anforderungen hinzugekommen, so dass auch diese durch das neue ELT-Tool abzudecken sind.

Wir werden im Folgenden abkürzend den Begriff „ELT“-Tool nutzen, wenngleich hierbei natürlich auch andere Verarbeitungsschritte bzw. Reihenfolgen mitgemeint sind.

Zur Überwindung der Herausforderungen, die durch die immer größer werdenden Datenmengen entstehen, muss ein ELT-Tool zur Datenbewirtschaftung **einfach** zu benutzen sein, damit Entwicklerinnen und Entwickler schnell und effektiv Prozesse ändern oder erweitern können.

Neben Entwicklungsumgebungen muss die Software auch eine **Serverkomponente** enthalten.

Zudem muss das Tool in die bestehende **Infrastruktur** des AOK-Bundesverbandes integrierbar sein und den Anforderungen an IT-Sicherheit und Datenschutz genügen.

Die Datenverarbeitungsprozesse sind derzeit im Wesentlichen als Kontrollflüsse angelegt. Das bedeutet, dass über Meta-Tabellen der Datenbank und über entsprechende Parameter der Prozessketten die Datenverarbeitung gesteuert wird. Dabei wird in der Regel durch **PUSH-Down** die eigentliche Datenverarbeitung auf den Datenbankservern durchgeführt. Es gibt aber auch Prozesse, die Datenflüsse in einen Kontrollfluss einbetten.

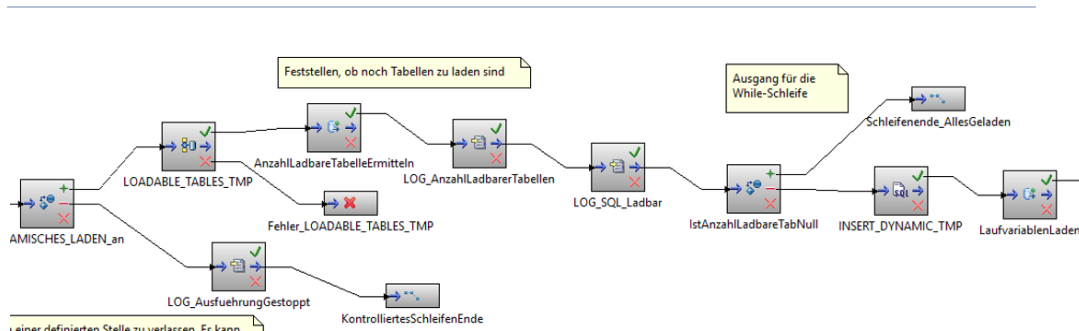
Eine **Ablaufsteuerung** bestimmt die Verarbeitungsreihenfolge von Datenflüssen, Data-Mining-Flüssen und anderen Arten allgemeiner Datenverarbeitungstasks (z. B. Ausführen von Betriebssystemskripten, SQL -Skripten, SFTP und E-Mail). Zudem ist es möglich iterative Schleifen, Bedingungen und Fehlerbehandlungen durchzuführen.

Eine typische Ablaufsteuerung ist zum Beispiel das Holen der Daten vom Filesystem, der Aufruf eines SQL-Skriptes zum Laden in eine Staging-Tabelle sowie die daran anschließenden Weiterverarbeitungen, bei denen ggf. neue Range Partitions für Tabellen erstellt und SQL- oder Shell-Skripte ausgeführt werden. Die Protokollierung in die Datenbank erfolgt direkt über einen entsprechenden SQL-Operator. Datenquellen liegen zum Teil als File vor und werden in der Regel über entsprechende Load-Aufrufe in die Datenbanken geladen. Auch Datenquellen anderer Datenbanksysteme als dem Zielsystem sollen verarbeitet werden können.

Zudem werden monatlich Katalogdaten aus dem Internet, die in XML-Format oder anderen gängigen Daten-Austausch-Formaten vorliegen, aus öffentlichen Quellen heruntergeladen und verarbeitet. Auch das Verarbeiten von formatierten Excel-Tabellen mit mehreren Tabellenblättern ist eine weitere Anforderung, da Statistikinformationen öffentlicher Quellen regelmäßig verarbeitet werden müssen.

Die Bausteine der gesamten Prozesskette werden in der Entwicklungsumgebung **visualisiert** dargestellt. Jeder einzelne Baustein kann separat inhaltlich definiert bzw. ausgefüllt werden.

Abbildung 1: Exemplarischer Auszug aus einem Prozess



2.1 Anforderungen an die Software

Die Anforderungen des Auftraggebers an die Softwarelösung sind im Detail im Anforderungskatalog (**Anhang zur Leistungsbeschreibung, „B3.1_Anforderungskatalog und Bewertungskriterien_*.xlsx“**) aufgeführt und mit eindeutigen Identifikatoren gekennzeichnet (Spalte „Anforderungsziffer“).

Die Einzelanforderungen sind in **Muss- und Kann-Kriterien** unterteilt, die im Anforderungskatalog beschrieben sind.

Hier eine Übersicht zu zentralen Anforderungen:

Die durch den Auftragnehmer zur Verfügung zu stellende Softwarelösung muss aus einer Entwicklungsumgebung und einer Serverkomponente bestehen.

Die allgemeinen Anforderungen an die Softwarelösung adressieren die intuitive Bedienbarkeit, zentrale Elemente einer Ablaufsteuerung sowie die Möglichkeit, dass Fachteams gemeinsam an Projekten arbeiten können. Relevant sind weiterhin die Systemintegration und der sichere Betrieb der Software.

In der Entwicklungskomponente werden die Datenverarbeitungsprozesse entwickelt und in eine Serverkomponente deployed. Die erstellten Prozesse werden damit auch nicht-Entwicklerinnen und -Entwicklern (Business User) zur Verfügung gestellt, die diese Prozesse über eine Weboberfläche aufrufen können. Dies gilt auch für entwickelte Modelle aus maschinellem Lernen (AUTO-ML) oder sonstigen statistischen Modellen. Im Folgenden eine Liste der zentralen Anforderungen an die Softwarelösung des Auftragnehmers:

- Vollständiger **On-Premises-Betrieb** (Idealerweise VM).
- Windows oder SUSE-LINUX
- Es können vielfältige Input-Datenformate verarbeitet werden.
 - Relationale Datenbanken -- **Datenbankunabhängigkeit** (z. B. PostgreSQL, Oracle, DB2)
 - Dateibasierte Verarbeitung (CSV, Excel, JSON etc.)
 - APIs / Webservices stehen zur Verfügung
- **DSGVO**-konformer Betrieb.
- Übersichtliches und gut handhabbares, rollenbasiertes **Berechtigungskonzept** und eine wirksame Zugriffskontrolle sind relevant. (Admin, Developer, Business User).
- Single-Sign-On (LDAP, Active Directory, SAML).
- Verschlüsselte Kommunikation (TLS/SSL).
- Möglichkeit zur Integration in bestehende Netzwerksicherheitsstrukturen (z. B. Firewall-Segmente, SSH-Tunnels).
- Updates und Security Patches müssen **On-Premises** installierbar sein.
- Unterstützung **kollaborativer** Workflows (Team-basierte Projekte), Versionsverwaltung
- Vollständig **grafische** Low-Code-**Oberfläche** (Drag & Drop) für ELT und Maschinelles Lernen. Es steht eine **Operatoren-Palette** zur Verfügung die per Drag and Drop bei der Entwicklung von Prozessketten genutzt werden können (No-Code-Prinzip).
- **Push-Down Processing**:
 - SQL-Transformationen sollen nach Möglichkeit direkt in der Datenbank ausgeführt werden.
 - Unterstützung von Spark-/Hadoop-Flows (falls verfügbar).
- Möglichkeit zur Erstellung wiederverwendbarer Komponenten.
- Eine Aufgabenplanung wird in der Serverkomponente angeboten (scheduling)
- Offene APIs (REST, JDBC/ODBC, Python/R SDKs) zur Integration von Programmierschnittstellen.
- Automated Machine Learning (**AutoML**) wird ebenfalls angeboten.
- Datenqualitätsfunktionen (wie Anomalieerkennung).
- Skalierbarkeit und Performance.
- Der Auftragnehmer muss Enterprise-Support bereitstellen (DE/EU bevorzugt).
- Dokumentation und Schulungsmaterialien müssen verfügbar sein.

On-Premises Lösungen

Auf Grund des sehr hohen Schutzbedarfs der zugrundeliegenden Daten sowie der erforderlichen Kontrolle über den Betrieb der eingesetzten Software durch den Auftraggeber darf der Auftragnehmer nur **On-Premises Lösungen** einsetzen.

Die Speicherung und Verarbeitung von Daten auf externen Servern, wie es bei SaaS-Lösungen üblich ist, ist aus rechtlichen und sicherheitstechnischen Gründen nicht zulässig. Dies vor allem vor dem Hintergrund, dass mit dem ELT-Tool auch Rohdaten verarbeitet werden. Somit werden teilweise sensible Daten verarbeitet, die den gesetzlichen Anforderungen des Sozialgesetzbuches unterliegen. Diese Daten mit dem **Schutzbedarf „sehr hoch“** können u.a. den Sozialdaten gleichgestellte Betriebs- und

Geschäftsgeheimnisse der Geschäftspartner der AOKs gemäß § 35 Abs. 4 SGB I sein. Daher sind On-Premises-Lösungen zu bevorzugen. Die Software wird als zentrales Datenbewirtschaftungstool eingesetzt, welches eng mit internen und betriebskritischen Prozessen verzahnt ist. Dies erfordert eine hohe Verfügbarkeit und **unmittelbare Kontrolle des Auftraggebers über alle Systemkomponenten**. Damit verbunden ist unter anderem eine granulare Kontrolle über Updates, Verfügbarkeit und Performance. Dies ist nur bei einer On-Premises-Lösung möglich.

Beständigkeit und Wartung der angebotenen Software

In der **Entwicklungsperspektive der nächsten 6 Jahre** muss nach jetzigem Stand sichergestellt sein, dass die **On-Premises Lösung** weiterhin zur Verfügung stehen wird. Dies vor dem Hintergrund, dass die Implementierung von Datenbewirtschaftungsprozessen mit entsprechenden Aufwänden des Auftraggebers verbunden ist.

Der Auftragnehmer muss stets sicherstellen, dass etwaige Sicherheitslücken der Softwarelösung schnell gefixt werden und entsprechende Softwarepatches zur Verfügung gestellt werden.

Datenverarbeitung über Kontrollflüsse in den zugrundeliegenden Datenbanksystemen

Die über die Software angesteuerte Datenverarbeitung muss primär in den zugrundeliegenden leistungsstarken Datenbanken bzw. Quellsystemen und darf nur teilweise innerhalb des Tools erfolgen. Lösungen, bei denen die Verarbeitungen bzw. Transformationen ausschließlich in ihrer eigenen Umgebung durchgeführt werden können, sind ausgeschlossen, da damit weitere Investitionskosten des Auftraggebers in eine geeignete Umgebung notwendig wären.

Deployment und Serverkomponente

Die entwickelten Flüsse/Prozesse werden auf einen Server deployed und stehen über eine Weboberfläche zur Verfügung. Insgesamt sollte der Deployment Prozess intuitiv sein und mit wenigen „Klicks“ möglich sein. Die deployten Prozesse müssen auf der Serverkomponente übersichtlich und mit einem Kommentarfeld dargestellt werden können. Der Filename des deployten Prozesses kann durch einen entsprechend sprechenden Namen ersetzt werden.

Anforderungen Maschinelles Lernen und Statistik.

Es besteht die Anforderung, Methoden des maschinellen Lernens über AUTO-ML Methoden anzuwenden, so dass die Daten mit minimalem Aufwand und minimalen Fachkenntnissen analysiert werden können. Hierzu muss es möglich sein, dass Trainingsdaten vorbereitet werden, entsprechende Modelle trainiert werden und anschließend die Modelle bewertet und ggf. iteriert werden. Die Schlussfolgerungen aus den Modellen sollen dann im Anschluss verglichen werden. Wichtig sind hierbei **Klassifizierungsmodelle, Regressionsmodelle** und **Prognosemodelle**. Hierbei werden Lösungen, die Datenvisualisierung und AUTO-ML Modelle anbieten, bevorzugt.

Damit verbunden soll auch automatisches **Feature Engineering** möglich sein. Mit automatischem **Feature Engineering** können aus den Daten die für die jeweiligen Modelle relevanten Informationen hervorgehoben werden und um irrelevante Merkmale reduziert werden. Somit wird insgesamt die jeweilige Modellkomplexität reduziert und die Interpretierbarkeit erhöht. Transparenz und Erklärbarkeit der Modelle sind ebenfalls relevante Features zur Modellinterpretierbarkeit mit Hilfe von Feature Importance oder SHAP (SHapley Additive exPlanations).

Auch spielen Themen zur Analyse von neuen Eingangsdaten eine Rolle. Hierbei geht es im Prinzip um **deskriptive statistische Auswertungen**. **Visualisierungen** der Eingangsdaten wie zum Beispiel BoxPlots, Histogramme etc., sowie die Darstellung von Ausreißern und Anomalie-Detection sind ebenfalls von Interesse.

2.2 Beispiele als Ergänzung zum Kriterienkatalog

2.2.1 Beispiel zu Deployen Teil 1 Muss-4.1.1:

Der ELT-Prozess selbst heißt zum Beispiel File2Mart_EXPORT_RKI_COVID. Dieser hat als Beschreibung den Text: „RKI-Daten Export für W18“. Beim deployen wurde der Name EXPORT_COVID_RKI vergeben.

Abbildung 2: Beispiel Prozesse in der Serverkomponente Deployment: Anwendungsname und Kommentar/Bemerkungsfeld

Anwendungen verwalten | Steuerungsflüsse verwalten | Instanzen verwalten | Zeitpläne verwalten

Die implementierten Anwendungen werden unten aufgelistet. Sie können eine neue Anwendung oder Änderungen an einer vorhandenen Anwendung implementieren.

Anwendungsname	Status	Beschreibung
FILE2MART_EXPORT_RKI_COVID	◆	RKI-Daten Export für W18
FILE2MART_LOAD_ADJ_PROD	◆	
FILE2MART_LOAD_FROM_CSV	◆	Laden von CSV-Dateien mit Spaltenbezeichnung in der ersten Zeile
FILE2MART_LOAD_VER_PROFIL_DATEN	◆	Tabelle VER_PROFIL_DATEN wird aus den einzelnen VER_XX-Tabellen befüllt

Anwendungen verwalten | **Steuerungsflüsse verwalten** | Instanzen verwalten | Zeitpläne verwalten

Die Steuerungsflüsse aus implementierten Anwendungen werden unten aufgelistet. Sie können einen Steuerungsfluss auswählen und ihn jetzt ausführen oder einen Zeitplan für ihn erstellen.

Steuerungsflussname	Anwendungsname	Beschreibung
EXPORT_RKI_COVID	FILE2MART_EXPORT_RKI_COVID	
LOAD_ADJ	FILE2MART_LOAD_ADJ_PROD	
LOAD_CSV	FILE2MART_LOAD_FROM_CSV	

2.2.2 Beispiel zu Protokollierung/Monitoring Teil 1 Muss-4.2.2

Der ELT-Prozess selbst heißt zum Beispiel Storage2MartV1. Beim deployen wurde der Name Storage2MartV1 vergeben. Ein Job der nun Daten verarbeitet wird mit den entsprechenden Parametern gestartet und wurde mit VER_QRT_ALL benannt. Der Job hat einen Status: Grün.

Abbildung 3: Beispiel Steuerungsfluss-Jobs mit Status (grün = ok, rot=Fehler)

Auf dieser Seite werden die Steuerungsflussinstanzen angezeigt, die aktiv sind oder deren Ausführung abgeschlossen ist. Die Liste der Steuerungsflussinstanzen auf dieser Seite kann erneut erstellt werden. Abhängig von der Anzahl Steuerungsflussinstanzen kann das erne Erstellen der Liste einige Zeit dauern.

Instanzname	Steuerungsfluss	Anwendungsname	Status	Startzeit	Abgelaufene Zeit
StopStorageStartLoading_A	StopStorageStartLoading	StopStorageStart	✓	22.07.2025 17:28:58	0:101s
VER_ORT_ALL_TEIL2	StorageStart/1	StorageStart/1	✓	22.07.2025 17:23:37	03m32.120s
VER_180_DMP_ADIPOSTAS_BASISMENGE_PM1	StorageStart/1	StorageStart/1	✓	22.07.2025 16:50:41	07m11.882s
EPLEPSE_VRM_10	LOAD_ADU	FILEMART_LOAD_ADU_PROD	✓	22.07.2025 16:48:02	08m34.807s
VER_DMP_ADIPOSTAS_10	StorageStart/1	StorageStart/1	✓	22.07.2025 16:40:22	07m11.214s
ADJ_EPLEPSE_VRM_A	LOAD_ADU	LOAD_BUR_HZ	✗	22.07.2025 16:02:23	02m06.603s
VER_ORT_ALL	StorageStart/1	StorageStart/1	✓	19.07.2025 16:53:41	02m29.37m49.310s
ETA_202501	LoadProfileFile	LoadProfileFile	✓	18.07.2025 16:43:57	08m59m01.464s
VER_180_AUTOMIBUL_P2	LOAD_VER_PROFIL_DATEN	FILEMART_LOAD_VER_PROFIL_DATEN	✓	17.07.2025 16:00:30	03m39.845s
VER_180_AUTOMIBUL	StorageStart/1	StorageStart/1	✓	17.07.2025 16:42:03	13m1.570s

2.2.3 Beispiel zu Logging und Monitoring Teil 1 Muss-4.2.1

Im Logging von Prozessen müssen auch Ausgaben aus den aufgerufenen Skripten ersichtlich sein. So wurde zum Beispiel ein Shell-Skript ausgeführt, welches über ECHO-Befehle bestimmte Ausgaben erzeugt. Diese werden dann auch im Logging ausgewiesen.

Abbildung 4: Beispiel eines Protokolls bei dem die Inhalte der Jobausführung eines Shell-Skriptes angezeigt werden

		<pre>SQW03258I: Execution result for DB2_LadenMitScript is PREPARED.VER_177_DMP_EFF_ASTHMA ===== Database Connection Information Database server = DB2/LINUX8664 11.5.8.0 SQL authorization ID = DB2 Local database alias = PRODDB ===== nachfolgende Fehlermeldung nicht schlimm da Tabs ggf schon geloescht ===== Return Status = 0 Return Status = 0 DB20000I The SQL command completed successfully. ===== A/IB_ICD_177 A/IBULANT ===== DB20000I The SQL command completed successfully. ===== SUBID = 0 nur gesichert SUBID = 1 Zustand Nach Diagnosen nur J46 als Z ===== Number of rows affected : 4368654 DB20000I The SQL command completed successfully. DB20000I The RUNSTATS command completed successfully. ===== ERGEBNIS ATC_DDD_177 A/IBULANT ===== DB20000I The SQL command completed successfully. DB20000I The SQL command completed successfully. DB20000I The SQL command completed successfully. Number of rows affected : 923832 DB20000I The SQL command completed successfully. =====</pre>
FINEST	27.08.2023 21:17:08	

3 Nutzung der Software:

Die Software wird zum einen von Personen verwendet, welche die Prozesse entwickeln (Entwicklerinnen und Entwickler), Personen, die User/Usergruppen anlegen und selektiv berechtigen, Datenbankverbindungen einrichten und Projekte anlegt usw. (Admins), sowie zum anderen von Personen, die über die Serverkomponente die Prozesse starten (Business User). Business User brauchen über das Starten von Prozessen hinaus keine weiteren Rechte zur Entwicklung oder Administration.

Die Entwickler-, Admin- und Business User-Lizenzen werden nach Usern (Named oder Concurrent), die Serverlizenzen gerätebasiert gezählt.

Die untenstehende Tabelle führt die Anzahl der benötigten Nutzenden auf. Je nach Lizenzmodell können sich mehrere Personen eine Lizenz teilen. Die Anzahl der Personen die gleichzeitig arbeiten können müssen ist in der Spalte „gleichzeitige Nutzungsmöglichkeit“ angegeben.

Zu Beginn des Vertrages rechnet der Auftraggeber mit folgenden Personen:

Tabelle 1: Anzahl Nutzenden zu Beginn des Vertrages

Nutzerkategorie	Anzahl der Nutzer	Gleichzeitige Nutzungsmöglichkeit
Entwickler	15	5
Admin*	1	1
Server	1	1
Business User**	30	10

*sofern abweichend von Entwicklern, sonst inkludiert in Entwickler

**sofern zusätzlich zur Serverkomponenten erforderlich

© WIdO 2026

Als maximalen Bedarf beziffert der Auftraggeber folgenden Personen:

Tabelle 2: Anzahl Nutzenden maximaler Bedarf

Nutzerkategorie	Anzahl der Nutzer	Gleichzeitige Nutzungsmöglichkeit
Entwickler	24	8
Admin*	2	1
Server	2	2
Business User**	100	35

*sofern abweichend von Entwicklern, sonst inkludiert in Entwickler

**sofern zusätzlich zur Serverkomponenten erforderlich

© WIdO 2026

4 Installations- und Betriebsbedingungen für IT-Systeme im AOK-Bundesverband

Die IT-Infrastruktur des AOK-Bundesverbandes definiert bestimmte Anforderung hinsichtlich Integration, Datenschutz, IT-Sicherheit und Wartbarkeit von Softwareprodukten. Die folgenden Anforderungen sind durch das vom Auftragnehmer eingesetzte ELT-Tool zwingend zu erfüllen. Diese Kriterien sind auch im **Anforderungskatalog (Anhang 1 zur Leistungsbeschreibung)** aufgeführt.

Der Auftragnehmer hat im Rahmen des Vergabeverfahrens Mindest-Hard- und Softwareanforderungen an die Entwicklungsumgebung sowie Serverumgebung in seinem Angebot zu benennen. Diese müssen sich in die folgenden Rahmenbedingungen des Auftraggebers einfügen:

4.1 Entwicklungsumgebung

- Betriebssystem: Installation der Entwicklungsumgebung ist auf mehreren Terminalservern (auf virtuellen Maschinen VMware) unter **Windows Server (ab Windows Server 2022)** vorgesehen.
- Verbindung zum **Internet** ist nur über einen **Proxy** mit einer **Kerberos** Authentifizierung möglich.
- **Den SSL-Zertifikaten** von Windows muss vertraut werden. Insbesondere den Root Zertifikaten die im Windows Truststore liegen.

4.2 Serverkomponente

Für die Serverkomponenten kann Linux **oder** Windows durch den Auftragnehmer eingesetzt werden. **Je nach Betriebssystem gelten die folgenden Anforderungen:**

- **Linux:** Als Linux Distribution kommt im AOK-Bundesverband nur **SUSE Linux Enterprise Server** zum Einsatz. Eine entsprechende Kompatibilität ist zu gewährleisten. Es wird mindestens SUSE Linux Enterprise Server 15 SP7 erwartet.
- Wenn der Server eine dedizierte Webserverkomponente benötigt (wie zum Beispiel Apache, nginx etc.) so müssen diese zwingend aus den Paketquellen der Suse-Distribution installierbar sein.
- **Windows:** Es wird mindestens Windows – Server 2022 erwartet.
- **Softwarebereitstellung über Containerlösung:** Im AOK-Bundesverband ist als Containerlösung Docker vorgesehen. Insofern die Software der Serverkomponente in einem Docker Container zur Verfügung gestellt wird, muss die Netzwerkkommunikation zwischen Containern und Kontrollmöglichkeiten für die zu öffnenden Ports gegeben sein. Die Kommunikation zwischen den Containern muss über lokale Firewall, Zertifikate und SSL abgesichert werden können. Auch die Containersoftware muss regelmäßig (mehrmals pro Jahr) durch den Auftragnehmer auf Sicherheitslücken geprüft werden. Bei Bedarf hat der Auftragnehmer Patch-Updates

zu veröffentlichen. Es sind auch die Verfahren zur Sicherstellung der Docker Image Security der bereitgestellten Images zu beschreiben.

4.3 Softwareanforderungen

- Falls Java verwendet wird ist folgender Releasestand zu berücksichtigen: Java Runtime, OpenJDK Java 22 JRE oder höher
- Die Kompatibilität der Webbrowser mit Microsoft Edge ab Version 135 muss gegeben sein
- Wird für den Webserver (Applikationsserver) Tomcat verwendet, so wird mindestens Tomcat 9.x oder höher erwartet.

4.4 Anforderungen an IT- und Datensicherheit

Prinzipiell besteht hinsichtlich der Softwarelösung eine hohe Anforderung an die Sicherheit. Dies bezieht sich sowohl auf die Software selbst als auch auf die Datenbankzugriffe sowie Zugriffe auf das Filesystem.

Hierzu zählen unter anderem folgende Anforderungen:

Generelle Anforderungen

- **Passwort**-Verschlüsselung (für User, der sich zur Zieldatenbank / zum Zielserver verbindet) muss gegeben sein
- Der Auftragnehmer hat den Auftraggeber über eventuell auftretende **Sicherheitslücken** im Produkt zu informieren und zeitnah entsprechende Patches zur Verfügung zu stellen

Hierzu zählen:

- Erfassung von Schwachstellen
- Bewertung zur Ausnutzbarkeit
- Behebung von Schwachstellen und Bereitstellen von Patches durch den Softwareanbieter
- **Datenbankverbindung** erfolgt durch JDBC oder ODBC - Schnittstellen
- Die Übertragung und Speicherung der Daten werden **nach dem aktuellsten Stand der Technik** verschlüsselt.

4.5 Mandantentrennung

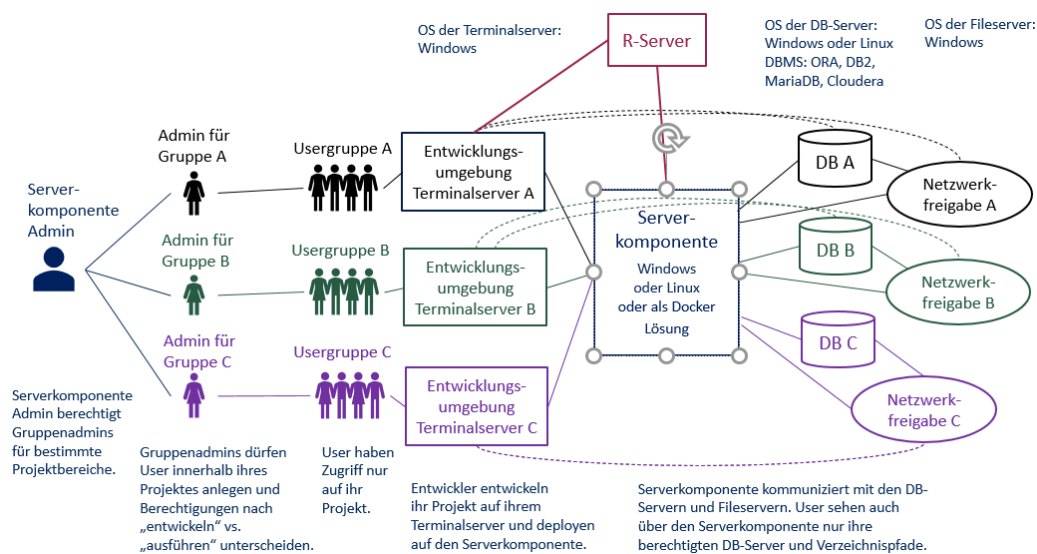
Für das Rollen- und Berechtigungskonzept erwarten wir eine Trennung von Berechtigungen je Forschungsbereich (Mandant). Somit können wir steuern, dass Personen des einen Forschungsbereichs nicht auf Prozesse oder Datenbankverbindungen anderer Forschungsbereiche zugreifen können.

Da das Thema **Mandantentrennung** von Bedeutung ist, wird zusätzlich je Forschungsbereich (Mandant) ein entsprechender Administrator festgelegt, der die Berechtigungen für das jeweilige Team definieren kann. Diese dürfen selbstredend nicht die Berechtigungen untereinander festlegen.

Über ein dediziertes Berechtigungskonzept muss über einen Administrator eingestellt werden können, wer welche Prozesse deployen oder ausführen kann. Das Abkapseln von Prozessen, Projekten und Datenbank-Verbindungen erfolgt nach Nutzendengruppen bzw. Mandanten

Siehe auch exemplarisch die untenstehende Abbildung.

Abbildung 5: Grafische Darstellung des Berechtigungskonzepts



Somit muss also gegeben sein, dass Prozesse nur für berechtigte Personengruppen sichtbar sind. So kann zum Beispiel Usergruppe A nicht die Prozesse von Usergruppe B einsehen.

5 Unterstützung bei der Inbetriebnahme

Für die Inbetriebnahme der Softwarelösung benötigt der Auftraggeber im Einstieg Unterstützung durch den Auftragnehmer. Dies erfolgt in Form von Dienstleisterunterstützung in Zusammenarbeit mit unsere Hauseigenen IT und den durch diese vorgegebenen Datenschutzregelungen. Entsprechende Zugänge werden bereitgestellt.

- Aufsetzen der Serverkomponente.
- Einrichten der Software für Entwickler auf VMs
- Einrichten der Lizenzen und sonstige Konfigurationen die zur Inbetriebnahme notwendig sind.
- Einführung zur Wartung und Update-Show-Case.
- Falls sich Software Updates paketieren lassen und damit ein automatisiertes Update der Clients (Entwickler) erfolgen kann, Unterstützung bei der implementieren dieses Szenarios.

Abbildungsverzeichnis

Abbildung 1: Exemplarischer Auszug aus einem Prozess	7
Abbildung 2: Beispiel Prozesse in der Serverkomponente Deployment: Anwendungsname und Kommentar/Bemerkungsfeld	10
Abbildung 3: Beispiel Steuerungsfluss-Jobs mit Status (grün = ok, rot=Fehler)	11
Abbildung 4: Beispiel eines Protokolls bei dem die Inhalte der Jobausführung eines Shell- Skriptes angezeigt werden.....	11
Abbildung 5: Grafische Darstellung des Berechtigungskonzepts	15

Tabellenverzeichnis

Tabelle 1: Anzahl Nutzenden zu Beginn des Vertrages	12
Tabelle 2: Anzahl Nutzenden maximaler Bedarf	12