

Leistungsbeschreibung **Data Mesh für die Olympiastützpunkte in Deutschland**

1. Ausgangslage und Handlungsbedarf

An den Olympiastützpunkten in Deutschland werden Daten aus zahlreichen Quellen erhoben und verarbeitet, etwa aus Mess- und Diagnostiksystemen, Formularen, Tabellen, Kalendern, Fachanwendungen, Anbieter-Clouds und lokalen Dateien. Die Daten liegen in unterschiedlichen offenen und proprietären Formaten und Semantiken vor und werden häufig in standortbezogenen Einzellösungen oder bei einzelnen Anbietern gespeichert.

Datenflüsse von Mess- oder Erfassungsinstanzen in weiterverarbeitbare Formate erfolgen vielfach manuell oder nur teilautomatisiert. Dadurch werden knappe personelle Ressourcen gebunden und Datenbestände verbleiben in Datensilos. Der Austausch mit anderen Olympiastützpunkten (OSP), IAT/FES, Verbänden, Trainerinnen und Trainern, Athletinnen und Athleten sowie weiteren Akteuren ist überwiegend in Einzelfalllösungen organisiert. Datenschutz, Zugriffskontrolle und Einwilligungs- bzw. Freigabemanagement werden dabei häufig nicht durchgängig technisch unterstützt.

Vor diesem Hintergrund soll eine gemeinsame technische Grundlage geschaffen werden, die die lokale Datenhoheit erhält und zugleich Interoperabilität, standardisierte Datenflüsse, kontrollierte Freigaben und eine bessere Auswertbarkeit ermöglicht.

2. Zentraler Usecase

Jeder OSP soll einen eigenen Data-Node mit eigener lokaler Datenhaltung betreiben können. Daten aus heterogenen bestehenden oder zukünftigen Datenquellen werden über wiederverwendbare Connector- und Mapping-Definitionen aufgenommen, geprüft, auf dokumentierte offene Ziel- bzw. Austauschstrukturen abgebildet und in dieser Datenbank gespeichert. Personenbezogene Fachdaten werden dabei mit einem Personen-/Athletenkern verknüpft.

Für die weitere Nutzung stellen berechnete Nutzerinnen und Nutzer aus freigegebenen Datenstrukturen wiederverwendbare und kontrollierte Datenprodukte zusammen. Diese werden insbesondere als Views, Exporte oder API-Zugriffe bereitgestellt und können von anderen Nodes sowie von BI- und Analysewerkzeugen genutzt werden. Aktualisierungen müssen abhängig von der jeweiligen Quelle wiederholbar und soweit technisch möglich automatisiert erfolgen.

Die Lösung muss so erweiterbar sein, dass weitere Datenquellen ohne grundlegende Änderung der Gesamtarchitektur integriert werden können. Der Nachweis erfolgt in der Referenzumsetzung anhand von mindestens zwei konkreten CSV-Formaten und einer konkreten externen Datenquelle über eine API.

3. Projektziel

Ziel des Vorhabens ist die Entwicklung eines roll-out-fähigen Netzes aus autonomen OSP-Data-Nodes. Jeder Node dient als lokale Dateninstanz, in der Daten importiert, geprüft, gespeichert, ausgewertet und kontrolliert freigegeben und geteilt werden können. Zwischen Nodes sowie mit weiteren berechtigten Akteuren soll ein strukturierter und protokollierter Datenaustausch möglich sein.

Es ist eine auf die Anforderungen des Auftraggebers zugeschnittene Individualsoftware als reproduzierbar installierbares und langfristig weiterentwickelbares System zu entwickeln. Das Vorhaben soll nicht nur abstrakt die Machbarkeit eines Data Mesh demonstrieren, sondern eine praktisch nutzbare Referenzversion für einen anschließenden Roll-out bereitstellen.

4. Kernziele

Dezentrale, souveräne Datenhaltung: Es wird keine zentrale Datensammelstelle geschaffen. Die Datenhoheit verbleibt beim jeweiligen Node bzw. der jeweils verantwortlichen Organisation. Personenbezogene Daten werden nur verarbeitet oder weitergegeben, wenn die erforderlichen technischen Berechtigungen sowie eine dokumentierte Freigabe, Einwilligung oder sonstige tragfähige Grundlage für den konkreten Zweck vorliegen.

Maximale Offenheit und Anschlussfähigkeit: Die Architektur beruht auf dokumentierten Schnittstellen, offenen Exportformaten und einer offenen Zugriffsschicht. Als Referenztechnologien dienen PostgreSQL oder eine gleichwertige Datenbanktechnologie, eine standardisierte Containertechnologie wie Docker oder eine gleichwertige Lösung sowie Apache Superset oder eine gleichwertige BI-Lösung. Andere BI- und Analysewerkzeuge müssen über dieselbe Zugriffsschicht angebunden werden können.

Praktisch nutzbare Entwicklung: Die Lösung wird als roll-out-fähiges System geliefert, das heterogene Datenquellen benutzerfreundlich in eine strukturierte lokale Datenhaltung überführt und den kontrollierten, rechte- und freigabebasierten Datenaustausch praktisch ermöglicht.

Anbieterwechsel-Fähigkeit und Rechte: Die projektspezifischen Arbeitsergebnisse (insbesondere der Quellcode) werden technisch dokumentiert und rechtlich so übergeben, dass der Auftraggeber das System durch einen anderen qualifizierten Anbieter betreiben, prüfen, migrieren und weiterentwickeln lassen kann (kein Vendor-Lock-in). Der Auftraggeber kann sämtliche Nutzungs- und Bearbeitungsrechte ohne Einschränkung mit Dritten teilen oder an diese übertragen.

5. Verbindlichkeit und Gleichwertigkeit

Die Architekturvorgaben – insbesondere autonome Nodes, modularer Monolith, containerisierte und reproduzierbare Bereitstellung, lokale Datenhaltung, offene Schnittstellen und Anbieterwechsel-Fähigkeit – sowie die in Abschnitt 8 bezeichneten Mindest-Ergebnisse sind verbindliche Mindestanforderungen. Weitere mit „muss“ oder „darf nicht“ formulierte Vorgaben sind ebenfalls verbindlich. Als Beispiele, Referenztechnologien, Soll-Vorgaben oder vom Bieter auszugestaltende Konzeptbestandteile gekennzeichnete Inhalte legen die konkrete Umsetzung dagegen nicht abschließend fest.

Soweit konkrete Produkte oder Technologien genannt werden, dienen diese als technische Referenz und sind jeweils im Sinne von „oder gleichwertig“ zu verstehen. Der Bieter kann alternative Technologien anbieten, muss deren Gleichwertigkeit oder Bessergeeignetheit jedoch im Angebot nachvollziehbar darlegen. Maßgeblich sind insbesondere Funktionalität, Interoperabilität, Selbstbetreibbarkeit, Sicherheit, Migrationsfähigkeit, Lizenz- und Betriebskosten, Dokumentierbarkeit und Vermeidung eines Vendor-Lock-ins.

6. Architektur-Grundsätze und technische Mindestanforderungen

6.1 Node-Architektur und Deployment

Die Umsetzung erfolgt als modularer Monolith pro OSP-Node. Die Anwendung wird als standardisiert deploybares Gesamtsystem entwickelt und intern in klar abgegrenzte Funktionsbereiche gegliedert. Sie wird containerisiert bereitgestellt; Docker oder eine gleichwertige Technologie kann hierfür verwendet werden.

Das Deployment muss reproduzierbar und dokumentiert sein. Manuelle, nicht dokumentierte Anbieter-Schritte sind zu vermeiden. Dieselbe Anwendung muss grundsätzlich lokal beim OSP, bei einem beauftragten Dienstleister oder in einer geeigneten Cloud betrieben werden können.

Neben der erstmaligen Installation ist auch die Aktualisierung eines bestehenden Nodes einschließlich Datenbankmigration, Sicherung und erforderlichenfalls Rückkehr auf einen vorherigen Stand sicherzustellen.

6.2 Datenkern und lokaler Personen-/Athletenkern

PostgreSQL oder eine gleichwertige Datenbanktechnologie bildet den lokalen Datenkern. Interoperabilität entsteht nicht allein durch ein einheitliches Datenbankprodukt, sondern insbesondere durch dokumentierte Datenmodelle, Connectoren, Mappings, Views, Exportformate und APIs.

Je Node ist ein flacher lokaler Personen-/Athletenkern vorzusehen. Dieser umfasst mindestens Name, Vorname, Geburtsdatum, Sportart und Kaderstatus. Alle personenbezogenen Fachdaten müssen auf diesen lokalen Kern rückführbar sein. Eine offene und dokumentierte API zur Aktualisierung des lokalen Kerns durch externe lokale oder zentrale Stammdatensysteme ist Bestandteil der Lösung.

Für die Zuordnung importierter Daten und den Austausch zwischen Nodes ist ein nachvollziehbares Identitäts- und Mappingverfahren vorzusehen. Unsichere Zuordnungen dürfen nicht unkontrolliert automatisch zusammengeführt werden; sie müssen prüfbar und erforderlichenfalls einer manuellen Entscheidung zugänglich sein. Eine zentrale Personenstammdatenbank, eine zentrale Identitätsautorität oder eine verpflichtende globale Athleten-ID ist nicht Bestandteil des Zielbilds. Organisationsübergreifende Zuordnungen müssen über dokumentierte und kontrollierte Identitäts- und Mappingverfahren erfolgen.

6.3 Datenintegration und Erweiterbarkeit

Die Lösung muss mindestens Datei-basierte und API-basierte Datenquellen unterstützen. Die Datenintegration umfasst Upload bzw. Abruf, Vorschau, Mapping, Plausibilitäts- und Formatprüfung, Fehler- und Importprotokoll, Speicherung und Bereitstellung über die einheitliche Zugriffsschicht (Export, API und/oder Ähnliches).

Wiederverwendbare Connector- und Mapping-Definitionen enthalten insbesondere Eingangsformat, Feldzuordnung, Einheiten, Plausibilitätsregeln, Kompatibilitätsregeln, Identifikationshinweise, Datenkategorie sowie externe Freigabe- und Teilbarkeits-Definitionen. Eingehende Rohdaten und die wesentlichen Verarbeitungsschritte müssen nachvollziehbar und auditierbar bleiben. Austauschschemas, Connector- und Mapping-Definitionen, Datenprodukte und APIs sind versionierbar zu gestalten. Änderungen müssen dokumentiert werden und dürfen bestehende Datenflüsse nicht unkontrolliert beeinträchtigen.

6.4 Frontend für Datenprodukte, Node-to-Node- und externer Datenaustausch

Es ist ein Frontend zur Definition kontrollierter Datenausschnitte vorzusehen. Berechtigte Nutzerinnen und Nutzer können daraus wiederverwendbare Datenprodukte erzeugen, insbesondere Views oder gleichwertige Abfragestrukturen, CSV-/JSON-/XML-Exporte und API-Endpunkte. Die pro Nutzer verfügbaren Datenbestände und Felder werden durch Rolle, Berechtigung, Zweck, Datenkategorie und Freigabe- bzw. Consent-Regeln begrenzt.

Auf dieser Basis wird ein Node-to-Node-Datenaustausch ermöglicht. Dieser ist nicht als unkontrollierte Gesamtsynchronisation zu konzipieren, sondern jeder Node bleibt eigenständig. Daten werden nur über definierte Austauschformate, APIs, Exporte oder Peer-Exchange-Funktionen geteilt, wenn hierfür eine Berechtigung und eine fachlich sowie rechtlich tragfähige Freigabe bestehen.

Austauschvorgänge müssen inhaltlich (insbesondere zeilen- und spaltenorientiert), zeitlich begrenzt, zweckgebunden und protokollierbar sein. Ein Node-Katalog bzw. eine gleichwertige Discovery-Funktion für inhaltsunabhängige Konfigurationen ist vorzusehen, ohne dadurch eine zentrale Datensammelstelle zu schaffen. Jeder Node verfügt dazu über einen lokalen Metadatenkatalog, in dem mindestens Datenquellen, Datenstrukturen, Felder, Connectoren, Mappings und Datenprodukte einschließlich Version, Datenkategorie, Teilbarkeit usw. nachvollziehbar beschrieben und freigegeben werden können. Kataloginformationen können kontrolliert mit anderen Nodes geteilt werden.

6.5 BI- und Analyseanbindung

Apache Superset oder eine gleichwertige BI-Lösung wird als vorkonfigurierte Standard-BI mit zwei bis drei einfachen Referenz-Dashboards integriert. Die BI-Lösung darf keine exklusive oder proprietäre Zugriffsschicht bilden. Excel, Power BI, SAP Analytics Cloud, Tableau, Python und andere Werkzeuge müssen die dokumentierten Views, Exporte oder APIs ebenfalls nutzen können.

6.6 Rollen, Access und Consent

Daten dürfen nicht pauschal austauschbar sein, sondern die Freigabe muss mindestens nach Datenschutz und Empfängerrolle kontrollierbar sein. Nicht freigegebene Daten sind standardmäßig auszuschließen. Es muss die Option geben, außerhalb des Systems gegebenen Consent manuell zu bestimmen, ohne dass die betroffenen Dateninhaber das im System selbst vornehmen (zum Beispiel Einwilligungserklärung auf Papier liegt bereits vor). Dazu muss das Angebot ein Konzept für Rollen, Access/Consent usw. enthalten. Zugriffe, Freigaben, Änderungen, Widerrufe und Austauschvorgänge müssen nachvollziehbar protokolliert werden. Das Angebot muss darstellen, wie mit Widerruf, Löschung und sonstigem Umgang mit bereits übertragenen Daten technisch und organisatorisch umgegangen wird.

Zur Umsetzung dieses Konzept-Elements gehört ein benutzerfreundliches Frontend.

6.7 Sicherheit, Betriebsvorbereitung und Zugänglichkeit

Der Bieter legt mit dem Angebot ein auf die vorgeschlagene Lösung bezogenes Sicherheits-, Datenschutz- und Betriebskonzept vor. Dieses ist im Projekt zu konkretisieren, umzusetzen und zu dokumentieren. Es muss insbesondere Verschlüsselung bei der Übertragung, sichere Authentifizierung, Rollen und Rechte, Schlüsselverwaltung, Protokollierung, Trennung externer und interner Dienste, Backup und Wiederherstellung, Schwachstellenmanagement, Updatefähigkeit und Incident Response berücksichtigen. Fehlgeschlagene Importe, Verarbeitungs- und Austauschvorgänge dürfen nicht unbemerkt bleiben, sondern müssen sichtbar, protokolliert und bearbeitbar sein. Backup und Wiederherstellung sind praktisch zu testen.

7. Mindest-Ergebnisse und übergreifende Abnahmekriterien

Es müssen mit der vollständigen Referenzversion mindestens folgende Ergebnisse nachweisbar erreicht werden:

- Der OSP-Data-Node kann anhand der übergebenen Dokumentation reproduzierbar als containerisiertes System installiert, konfiguriert und gestartet werden. Die Reproduzierbarkeit wird durch eine Neuinstallation in einer leeren Referenzumgebung allein anhand der übergebenen Artefakte und Dokumentation und ohne undokumentierte Unterstützung des Auftragnehmers praktisch nachgewiesen.
- Mindestens zwei konkrete CSV-Datenquellen und eine konkrete API-Datenquelle können jeweils durchgängig von der Aufnahme über Vorschau, Mapping, Prüfung und Speicherung bis zu View, Export bzw. API und BI verarbeitet werden.
- Der lokale Personen-/Athletenkern kann über eine dokumentierte API aktualisiert werden; importierte personenbezogene Fachdaten werden nachvollziehbar zugeordnet.
- Berechtigte Nutzerinnen und Nutzer können aus freigegebenen Datenstrukturen wiederverwendbare Datenprodukte erzeugen und als View oder gleichwertige Abfragestruktur, Export und API bereitstellen.
- Rollen-, Access- und Consent-/Freigaberegeln verhindern die Ausgabe nicht berechtigter Daten und protokollieren die maßgeblichen Entscheidungen.
- Das Frontend unterstützt benutzerfreundlich und in angemessener Responsivität auf mobile Endgeräte die geforderten und konzipierten Workflows, insbesondere Integration von Datenquellen, Export/API, Freigabe/Access/Consent und Administration, und zeigt Fehler sowie Verarbeitungsstatus nachvollziehbar an.
- Mindestens drei Node-Instanzen werden an Pilotstandorten oder in einer gleichwertigen Referenzumgebung eingerichtet; zwischen den Nodes wird gemäß der Anforderungen ein kontrollierter und protokollierter Datenaustausch praktisch nachgewiesen.
- Apache Superset oder die angebotene gleichwertige BI-Lösung ist angebunden; mindestens zwei Referenz-Dashboards greifen über die vorgesehene Zugriffsschicht auf Daten zu, Datenveränderungen werden automatisiert synchronisiert.
- Sicherheits-, Backup- und Wiederherstellungsfunktionen sowie Fehlerbehandlung werden anhand von Prüffällen nachgewiesen.
- Quellcode, Repository, Datenbankschemata, Schnittstellen, Deployment, Tests und Dokumentation werden so übergeben, dass ein sachkundiger Dritter das System grundsätzlich installieren, prüfen und weiterentwickeln kann. Dies wird durch einen Gutachter attestiert, der intern oder extern vom Auftraggeber beauftragt wird.

Die Einzelheiten ergeben sich aus der mit dem Angebot eingereichten Produktkonzeption.

8. Testdaten, Referenzquellen und Mitwirkung

Zur Angebotserstellung ist keine produktive Anbindung an einen bestimmten Geräte- oder Cloudanbieter erforderlich. Bieter können geeignete eigene Beispieldaten oder Test-APIs zur Veranschaulichung ihrer Konzeption verwenden und müssen die zugrunde gelegten Annahmen offenlegen.

Die für Umsetzung und Abnahme maßgeblichen zwei CSV-Formate und die API-Datenquelle werden zu Projektbeginn durch den Auftraggeber aus bereitgestellten oder von ihm freigegebenen Referenzquellen festgelegt. Der Auftraggeber stellt die verfügbaren Testdaten, Dokumentationen und Zugänge bereit. Ist der Zugriff auf eine reale Drittanbieter-API nicht rechtzeitig möglich, kann im Einvernehmen eine funktional vergleichbare Test-API verwendet werden. Der Bieter stellt in seinem Angebot transparent dar, welche Mitwirkungsleistungen und Zugänge er benötigt.

9. Rechte an Arbeitsergebnissen und Anbieterwechsel-Fähigkeit

Ein zentrales Prinzip des Projekts ist digitale Souveränität. Der Auftraggeber und die beteiligten OSP dürfen nicht faktisch an einen bestimmten Anbieter gebunden sein und müssen das Produkt innerhalb des organisierten Sports weitergeben sowie durch Dritte betreiben und weiterentwickeln lassen können.

Deshalb müssen alle projektspezifischen Arbeitsergebnisse vollständig zugänglich gemacht werden. Hierzu gehören insbesondere Quellcode und Repository-Struktur, Datenbankschemata und Migrationsskripte, Schnittstellen- und API-Dokumentation, Deployment-Konfigurationen, Testfälle und Testdaten, automatisierte Tests, Architekturentscheidungen, Betriebs- und Installationsdokumentation sowie eine Übersicht der eingesetzten Drittkomponenten und Lizenzen. Referenz-Repository ist GitHub, vergleichbare Alternativen können angeboten werden.

Die Einzelheiten der Nutzungs-, Änderungs-, Weiterentwicklungs-, Betriebs- und Weitergaberechte sowie der Übergabe werden im Vertrag geregelt. Die Anbieterwechsel-Fähigkeit ist Bestandteil der Leistung und der Abnahme.