Print Get PDF

# Final Report 0 on the DZIF Project

| | |
|---|---|
| **Project number** | TI 12.201 |
| **Title of the project** | DZG-übergreifende Dateninfrastruktur |

**Project partner and project duration**

| **Project Coordination** | **Start of Term** | **End of Term** |
|---|---|---|
| Universität zu Köln | 01.01.2021 | 28.02.2026 |

| **Project Partners** | **Start of Term** | **End of Term** |
|---|---|---|
| Universität zu Köln | 01.01.2021 | 28.02.2026 |

## Chapter 1: Short report (will be published)

### a) Original task and scientific/technical status built on

#### Universität zu Köln

Ziel des Vorhabens war die Etablierung einer DZG-übergreifenden Dateninfrastruktur, die den Austausch von Daten, Werkzeugen, Expertise und Prozessen zwischen den Deutschen Zentren der Gesundheitsforschung (DZG) ermöglicht. Ausgangspunkt war die Beobachtung, dass die DZG zwar täglich umfangreiche, qualitativ hochwertige Daten aus klinischen Studien, Biomaterialanalysen und Grundlagenforschung erzeugen, diese Daten jedoch aufgrund heterogener IT-Systeme, verteilter Speicherung und fehlender gemeinsamer Authentifizierungs- und Transferinfrastrukturen kaum DZG-übergreifend nutzbar waren. Da viele der von den DZG erforschten Volkskrankheiten voneinander abhängen, war eine engere inhaltliche Zusammenarbeit nicht nur fachlich naheliegend, sondern Voraussetzung für die wissenschaftliche Konkurrenzfähigkeit der DZG im Bereich der translationalen Gesundheitsforschung.
Das Vorhaben verfolgte explizit nicht das Ziel der Harmonisierung um jeden Preis, sondern setzte auf Interoperabilität: bestehende, in Benutzung befindliche Lösungen einzelner DZG sollten evaluiert, angepasst und gegenseitig nutzbar gemacht werden. Die Aufgaben wurden zu Beginn der Förderperiode in sechs Arbeitspakete gegliedert:

1. AP1: DZG-Projektregister / DZG-Parameterverzeichnis
2. AP2: DZG-Datentransfer-Plattform
3. AP3: Data Science Orchestrator
4. AP4: Gemeinsame Authentifizierungsstruktur
5. AP5: Record Linkage
6. AP6: DZG-Infrastruktur / gemeinsame Außendarstellung

Wesentliche Abweichungen im Verlauf:

1. Institutioneller Umzug der DZIF-zuständigen AG Kohorten von der Uniklinik Köln zum IDMKD an der Goethe-Universität Frankfurt während der Vorhabenlaufzeit. Damit war die ursprünglich in Köln geplante DZIF-spezifische Implementierung einzelner Komponenten (insbesondere Datentransferplattform-Instanz) nicht mehr am vorgesehenen Standort umzusetzen; die Weiterführung erfolgt im Folge-Vorhaben TI 12.202 am IDMKD Frankfurt.
2. AP5 (Record Linkage) wurde nach erfolgreichem Machbarkeitsnachweis auf synthetischen Daten nicht in eine produktive DZG-übergreifende Lösung überführt, da kein konkreter Anwendungsfall identifiziert werden konnte (DZG-AG-Jahresberichte 2023, 2024). Die Schnittstellenanforderungen wurden in die NUM-Strukturen eingebracht.

### b) Course of the project

#### Universität zu Köln

Die Arbeit erfolgte in iterativen, eng abgestimmten Entwicklungszyklen mit zweiwöchentlichen Koordinationsmeetings der AG sowie wöchentlichem Entwickler-Jour-Fixe. Pro Jahr fanden ein bis zwei Präsenztreffen statt (u.a. Heidelberg 2022 beim DKFZ, Gießen Zwischenevaluation Oktober 2024, DZG-Tag Heidelberg 2024, München Juni 2025 zur Planung der Weiterführung der Arbeit im Folge-Vorhaben). Eine gemeinsame Cloud-Plattform (cloud.dzg.digital) wurde 2021 etabliert; sie wird seither durchgehend für AG-interne Zusammenarbeit produktiv genutzt und hat sich als DZG-AG-übergreifende Infrastruktur etabliert, die auch von weiteren DZG-Arbeitsgruppen außerhalb der AG Forschungs-IT für ihre Zusammenarbeit verwendet wird.
Die Arbeiten der AG Kohorten (DZIF-Anteil; bis zum Umzug an der Uniklinik Köln, danach am IDMKD Frankfurt) lagen schwerpunktmäßig auf:

1. AP4: DZIF-seitige Anbindung an die gemeinsame Authentifizierung,
2. AP1: Aufbau der Studienmetadaten-Pipeline und des Web-Frontends (in Kooperation mit NFDI4Health),
3. AP2: DZIF-seitige Einbindung in das DZG Data Sharing Framework,
4. AP6: DZIF-Gremienarbeit und Außendarstellung.

Bei AP3 und AP5 lag der Entwicklungs-Lead beim DKTK; die AG Kohorten beteiligte sich konzeptionell und mit gezielten Eigenleistungen (Containerexecutor-Hackathon, Reproduzierbarkeits-Komponente in AP3; Testdatensatz und experimentelle Beteiligung in AP5).

Im Laufe des Vorhabens wurde die Zusammenarbeit mit dem neu gegründeten DZPG (ab 2023 in der AG vertreten) sowie die Vernetzung mit externen Initiativen (NUM, NFDI4Health, MII, IAM4NFDI) substantiell intensiviert.

### c) Main results and cooperation with other research institutions, if applicable

#### Universität zu Köln

Wesentliche, am Ende der Laufzeit verfügbare Ergebnisse:

1. **DZG-Cloud-Plattform** (cloud.dzg.digital): Seit 2021 produktiv betriebene Nextcloud-basierte Kollaborationsplattform mit IT-Sicherheitskonzept und DSFA-Bewertung der UKK-Stabsstelle Datenschutz. Hat sich über die AG Forschungs-IT hinaus zur DZG-AG-übergreifend genutzten Infrastruktur entwickelt.
2. **DZG-übergreifender Basisdatensatz** (AP1): Konsentierung von etwa 100 Datenitems mit mandatierten Kliniker:innen aller DZG abgeschlossen, Erstveröffentlichung im MDM-Portal am 19.09.2023 (medical-data-models.org/45851, DOI 10.21961/mdm:45851, Lizenz CC BY-NC-SA 4.0), Codierung mit LOINC und SNOMED. Während der Vorhabenlaufzeit wurden zwei weitere aktualisierte Versionen veröffentlicht (12.07.2024, 28.03.2025). Verpflichtende Erhebung in DZG-finanzierten Studien durch DZG-Board 2023 beschlossen, Implementierung 2024–2025.
3. **DZG-Studienmetadatenkatalog** (AP1): In Kooperation mit NFDI4Health entwickelter Prototyp eines föderierten Studienregisters mit Web-Frontend (studies.dzg.digital), Datenpipeline und Web-Crawler für ClinicalTrials.gov und DRKS.de. Open-Source-Veröffentlichung.
4. **DZG Data Sharing Service** (AP2): Auf Nextcloud basierender, dokumentierter und 2024 abgeschlossener Dienst zum sicheren Datentransfer (git.apps.dzd-ev.org/dzdtools/dzddatasharingservice).
5. **Data Science Orchestrator (DSO)** (AP3): Im Lead des DKTK entwickelter Prototyp einer Trusted Research Environment mit containerisierten, modularen Komponenten (Container Factory, Management Console, Trusted Partition, Secure Interaction Console, Datalad-basierte Versionierung, Reproducible Execution Environment, Data Outlet Review). Co-Autorenschaft aus der AG Kohorten (Brechtel) an der gemeinsamen MIE-2025-Publikation: Kern J, Brechtel M, Kaluza P, Kussel T, Vehreschild JJ, Lablans M. *Data Science Orchestrator: A Containerized Trusted Research Environment for Flexible and Secure Analytical Pipelines.* Studies in Health Technology and Informatics, Vol. 327 (Intelligent Health Systems – From Technology to Data and Knowledge), IOS Press 2025, S. 698–702. DOI: 10.3233/SHTI250438. Open Access (CC BY-NC 4.0).
6. **Föderierte DZG-Authentifizierung** (AP4): Produktiv eingerichtete OpenID-Connect-basierte föderierte Authentifizierung zwischen DZIF, DZD, DKTK und den zentralen DZG-Diensten unter dzg.digital. Definition organisationsübergreifender Benutzerrollen.
7. **Record-Linkage-Machbarkeitsstudie** (AP5): Erfolgreicher Test des DZG-übergreifenden Record Linkage auf synthetischen Daten 2022; Evaluation eines Verfahrens ohne Trusted Third Party (SMPC) und Bestätigung der technischen Durchführbarkeit. Im DZG-Kontext wurde kein konkreter Anwendungsfall identifiziert, der ein produktives DZG-übergreifendes Record Linkage erfordert hätte (DZG-AG-Jahresberichte 2023 und 2024). Die im Projekt gewonnenen Anforderungen an interoperable Record-Linkage-Schnittstellen wurden 2024 in die NUM-Strukturen eingebracht (DZG-AG-Jahresbericht 2024).
8. **Sichtbarkeit und Vernetzung** (AP6): Workshop auf der GMDS/TMF-Jahrestagung 2022, Teilnahme an der NUM Convention 2024, Artikel im DZG-Magazin Synergie (10/2024), Poster beim DZG-Tag 2024 in Heidelberg.

Kooperationen mit anderen Forschungseinrichtungen:

1. **NFDI4Health**: Kooperationsprojekt zur Entwicklung des Studienmetadatenkatalogs und der bidirektionalen Schnittstelle zum Study Hub.
2. **Netzwerk Universitätsmedizin (NUM)**: Beteiligung an Projektgruppen (u.a. zu ID- und

Consent-Management / Record Linkage) und an der NUM Convention 2024.
3. **DZPG**: Nachnutzung des DZG-Studienmetadatenkatalogs für das eigene Study/Cohort Repository; Konsentierung von Variablen zur psychischen Gesundheit für den erweiterten Basisdatensatz.
4. **TI BBD des DZIF** : Anschlussverwertung der Authentifizierungs- und Datentransferinfrastruktur.
5. **IAM4NFDI**: Abstimmung der Authentifizierungslösung mit den entstehenden nationalen Authentifizierungsstandards.
6. **Medizininformatik-Initiative (MII)**: Übernahme von Kerndatensätzen und FHIR-Profilen als Referenzen für den DZG-Basisdatensatz.

Die im Vorhaben entwickelten Lösungen werden ab 01.01.2026 im Folge-Vorhaben **TI 12.202 "DZG-übergreifende Dateninfrastruktur"** (Koordination: IDMKD Goethe-Universität Frankfurt) in produktive Systeme überführt und mit nationalen Infrastrukturen (NUM Hub, IAM4NFDI) integriert.

**Chapter 2: Detailed description (will be published)**

**a) Detailed presentation of the results achieved**

**Universität zu Köln**

# AP1: DZG-Projektregister / DZG-Parameterverzeichnis

**Plan laut VHB:** Aufbau eines DZG-übergreifenden Projektregisters und Parameterverzeichnisses zur Findbarkeit von Forschungsdatensätzen auf Metadaten-Ebene; schrittweise und freiwillige Standardisierung nach FAIR-Prinzipien.

**Erzielte Ergebnisse:**

1. 2021–2022: Konsentierung eines DZG-übergreifenden Basisdatensatzes mit mandatierten Kliniker:innen aller DZG; Definition eines Aktualisierungsprozesses.
2. 2023: Konsentierung von etwa 100 Datenitems für den erweiterten DZG-Datensatz abgeschlossen. Veröffentlichung des Basisdatensatzes im MDM-Portal am 19.09.2023 (medical-data-models.org/45851, DOI 10.21961/mdm:45851, Lizenz CC BY-NC-SA 4.0).
3. 2023: Beginn der durch NFDI4Health finanzierten Kooperation zur Entwicklung einer Website zur Darstellung des Metadatenkatalogs (DZG Studienübersicht) inklusive Web-Crawler für ClinicalTrials.gov und DRKS.de. Die AG Kohorten hat in einem Entwicklungsprojekt mit einem extern beauftragten Team eine Datenpipeline und das Web-Frontend zur Verarbeitung, Darstellung und Zusammenstellung der Studienmetadaten aller DZG entwickelt. Die Software wurde als Open-Source-Repository veröffentlicht.
4. 2024: DZG-Board beschloss die verpflichtende Erhebung des DZG-übergreifenden Basisdatensatzes in DZG-finanzierten Studien; Beginn der Umsetzung. Codierung des Basisdatensatzes mit LOINC und SNOMED zur interoperablen Nutzung. Studienmetadatenkatalog-Prototyp online unter studies.dzg.digital.
5. 2024: Erweiterter DZG-übergreifender Basisdatensatz mit mandatierten Kliniker:innen abgestimmt; Codierung in Vorbereitung.
6. 2025: Codierung des erweiterten Basisdatensatzes; Veröffentlichung von Folgeversionen im MDM-Portal (Version 2 am 12.07.2024, URL 46010; Version 3 am 28.03.2025, URL 46184). Insgesamt liegen am Vorhabenende drei während der Förderlaufzeit veröffentlichte Versionen des DZG-Basisdatensatzes öffentlich verfügbar vor (19.09.2023, 12.07.2024, 28.03.2025).

**Eigenanteil AG Kohorten:** Entwicklung und Bereitstellung des Web-Frontends (studies.dzg.digital), der Datenpipeline, der Filter- und Darstellungslogik sowie des Import-/Export-Mechanismus für die NFDI4Health-Study-Hub-Anbindung. Diese Arbeiten wurden mit Unterstützung externer Auftragnehmer durchgeführt (Finanzierung des Entwicklungsteams aus NFDI4Health-Mitteln).

# AP2: DZG-Datentransfer-Plattform

**Plan laut VHB:** Aufbau einer DZG-internen und DZG-übergreifenden Datentransfer-Plattform mit besonderer Berücksichtigung personenbezogener Daten nach EU-DSGVO; Einholen einer Empfehlung der TMF AG Datenschutz.

**Erzielte Ergebnisse:**

1. 2022: DSFA-Anfrage für die Datenaustauschplattform bei der Stabsstelle Datenschutz der Uniklinik Köln (UKK) gestellt; Vorstellung beim Datenschutz und der IT-Sicherheit der UKK. Die Bewertung durch Datenschutz und IT-Sicherheit der UKK fiel positiv aus.
2. 2023: Technische Konfiguration für einen sicheren Datentransferdienst auf Basis der datenschutzkonformen Open-Source-Lösung Nextcloud erarbeitet und dokumentiert (git.apps.dzd-ev.org/dzdtools/dzddatasharingservice).
3. 2024: Abschluss der Entwicklung des allgemeinen DZG Data Sharing Service.
4. DZIF-seitig: Das DZD-Data-Cloud-Konzept wurde dem Datenschutzbeauftragten und IT-Sicherheitsbeauftragten der UKK vorgestellt. Die produktive DZIF-Instanz der Datentransferplattform konnte innerhalb der Vorhabenlaufzeit nicht in Köln bereitgestellt werden, da die für die DZIF-IT zuständige AG Kohorten institutionell von der Uniklinik Köln zum IDMKD an der Goethe-Universität Frankfurt umgezogen ist. Damit war ein eigenständiger

Neuaufbau in Frankfurt erforderlich, der die ursprüngliche Köln-spezifische Implementierungsplanung obsolet machte.
5. Die produktive DZIF-Instanz der Datentransferplattform wird im Folge-Vorhaben TI 12.202 (AP4 Data Sharing Platform) am IDMKD Frankfurt neu aufgebaut. Parallel erfolgt die Anschlussverwertung im Rahmen der TI BBD.

**Eigenanteil AG Kohorten:** Vorbereitung, Vorstellung und Begleitung der Datenschutz-/IT-Sicherheits-Prüfung an der UKK; konzeptionelle Vorbereitung einer DZIF-Instanz, deren Bereitstellung durch den Umzug der AG Kohorten nach Frankfurt nicht abgeschlossen wurde. Die im Vorhaben erarbeiteten Anforderungen und Konzepte werden im Folge-Vorhaben TI 12.202 in Frankfurt umgesetzt.

# AP3: Data Science Orchestrator

**Plan laut VHB:** Prototyp eines Data Science Orchestrators (DSO), der mindestens ein Analysetool bereitstellt und Datenanalyse ohne direkten Datenzugriff über eine Web-Oberfläche ermöglicht.

**Erzielte Ergebnisse (Lead: DKTK):**

1. 2022: Konzept mit zwei Schwerpunkten (föderierte Ausführung von Analyseskripten; interaktive Analyseumgebungen in geschützten Umgebungen).
2. 2023: White Paper, Architektur- und Use-Case-Diagramm, Prototyp der Prozessierungsumgebung. Hackathon-Entwicklung eines Containerexecutor-Prototyps in Zusammenarbeit zwischen AG Kohorten und DKTK.
3. 2024–2025: Funktionsfähiger DSO-Prototyp als modulare, containerisierte Trusted-Research-Environment-Referenzarchitektur. Das System adressiert drei konsentierte Anwendungsfälle der DZG-Forschungs-IT: interaktive Analyse durch Forschende, automatisierte wiederkehrende Reporting-Pipelines und föderierte Multi-Standort-Analysen. Zentrale Komponenten umfassen eine Container Factory mit SBOM-Generierung (Syft, Trivy), Schwachstellen-Scanning (OSV-Scanner) und Malware-Prüfung; eine Management Console (PostgreSQL/FastAPI/Svelte) zur Datensatz- und Umgebungsverwaltung; Trusted Partitions auf Basis von Coder und OpenTofu/Terraform; eine Secure Interaction Console mit virtuellem Browser und VNC-Auditing; Datalad/Git-Annex-basierte Datenversionierung mit Trennung von internen und Export-Repositorien; ein Reproducible Execution Environment auf Basis von Concourse CI; und eine Data Outlet Review Console für manuelle und automatisierte Output-Reviews.
4. 2025: Open-Access-Publikation in Studies in Health Technology and Informatics (MIE 2025, IOS Press). Konzeption als Grundlage für ein DZG Trusted Research Environment im Hinblick auf Gesundheitsdatennutzungsgesetz (GDNG) und European Health Data Space (EHDS).

**Eigenanteil AG Kohorten:** Beteiligung an Konzept, Anforderungsanalyse und Architektur des DSO; gemeinsame Hackathon-Entwicklung des Containerexecutor; Beauftragung externer Entwicklung (Ghostroute Consulting, finanziert aus TI-12.201-Mitteln) für die Datenversionierungs- und Reproduzierbarkeits-Komponenten. Co-Autorenschaft an der MIE-2025-Publikation (Brechtel, Kaluza). Der Lead der Gesamtentwicklung lag beim DKTK; die AG Kohorten brachte strukturelle Beiträge zur Reproduzierbarkeit und zur Datenverwaltungs-Architektur ein.

# AP4: Gemeinsame Authentifizierungsstruktur

**Plan laut VHB:** Konzeptionelle Entwicklung einer DZG-weiten Authentifizierungsstruktur; Ergebnis kann ein Architekturkonzept, ein Optionenvergleich mit Handlungsempfehlung oder ein Pflichtenheft sein.

**Erzielte Ergebnisse:**

1. 2021: Konzept für eine gemeinsame DZG-Authentifizierung erstellt.
2. 2022: Prototyp einer föderierten OpenID-Connect-basierten Authentifizierung entwickelt.
3. 2023: Föderierte Authentifizierung zwischen DZIF, DZD, DKTK und den zentralen DZG-Diensten unter dzg.digital eingerichtet. Identifikation organisationsübergreifender Benutzerrollen. Anmeldung von DZG-Gastnutzer:innen am DZIF-Intranet ermöglicht.
4. 2024: Produktive Nutzung der Authentifizierungsinfrastruktur durch die AG Forschungs-IT. Definition der Schnittstellenanforderungen (OAuth2) für die Anbindung weiterer DZG.

Das in der VHB als Optionen genannte Spektrum (Architekturkonzept, Optionenvergleich mit Handlungsempfehlung oder Pflichtenheft) wurde im Vorhaben mit einer produktiv genutzten Referenzimplementierung umgesetzt.

**Eigenanteil AG Kohorten:** Technische Bereitstellung der DZIF-Schnittstellen; Beteiligung an der Rollendefinition; Integration in das DZIF-Intranet.

**AP5: Record Linkage**

**Plan laut VHB:** Experimentelles Record Linkage zur Beurteilung des Bedarfs und der Machbarkeit einer dauerhaften DZG-übergreifenden Pseudonymisierungslösung; Handlungsempfehlung an die DZG; angestrebte Publikation.

**Erzielte Ergebnisse (Lead: DKTK):**

1. 2022: Erstellung eines synthetischen Testdatensatzes durch die AG Kohorten; erfolgreicher Test des DZG-übergreifenden Record Linkage auf diesen Daten.
2. 2023: Evaluation einer Lösung ohne Trusted Third Party (SMPC-basiert) und Bestätigung der technischen Durchführbarkeit.
3. 2024: Im DZG-Kontext wurde kein konkreter Anwendungsfall identifiziert, der ein produktives DZG-übergreifendes Record Linkage erfordert hätte. Die im Projekt gewonnenen Erkenntnisse zur Heterogenität bestehender Record-Linkage-Lösungen und die daraus resultierenden Anforderungen an interoperable Schnittstellen wurden in die NUM-Strukturen eingebracht.

**Plan-Soll-Vergleich:** Die VHB definiert AP5 als *"experimentelles Record Linkage […], welches als Entscheidungsgrundlage dient, um den Bedarf einer dauerhaften Record Linkage/Pseudonymisierung in einem DZG-übergreifenden Kontext zu beurteilen"*. Die technische Machbarkeit wurde auf Testdatenbasis verifiziert (einschließlich SMPC-basierter Variante ohne Trusted Third Party); die Heterogenität bestehender Record-Linkage-Lösungen wurde dokumentiert. Im DZG-Kontext wurde während der Vorhabenlaufzeit kein konkreter Anwendungsfall identifiziert, der den Aufbau einer produktiven DZG-übergreifenden Pseudonymisierungsinfrastruktur erfordert hätte (DZG-AG-Jahresberichte 2023, 2024). Die im Projekt gewonnenen Schnittstellenanforderungen wurden in die NUM-Strukturen eingebracht (DZG-AG-Jahresbericht 2024).

Nicht erreicht wurde das in Meilenstein 7 formulierte Teilziel, einen DZG-übergreifenden Pseudonymisierungsdienst produktiv online zu stellen (Status "nicht erreichbar"; siehe Kapitel 2b).

**Eigenanteil AG Kohorten:** Erstellung des synthetischen Testdatensatzes; Beteiligung an den Experimenten; konzeptionelle Vorbereitung der späteren DZIF-seitigen Implementierung im Rahmen der TI BBD bzw. der NUM-Strukturen.

# AP6: DZG-Infrastruktur — gemeinsame Außendarstellung

**Plan laut VHB:** Vertretung der DZG in Gremien (MII, TMF, GBN); Stärkung der Sichtbarkeit als primäre Quellen translationaler Forschungsdaten.

**Erzielte Ergebnisse:**

1. 2021–2025: Jährliche Berichte über Gremienaktivitäten der Projektmitglieder.
2. 2022: GMDS/TMF-Workshop "Research IT Collaboration and Challenges across German Centres for Health Research (DZG)" auf der Jahrestagung in Köln.
3. 2022: Vorstellung der Vision und Mission der DZG-Forschungs-IT beim DZG Discussion Forum.
4. 2023: Veröffentlichung einer Auswahl gemeinsamer DZG-IT-Projekte auf der DZG-Website.
5. 2024: Teilnahme an der NUM Convention 2024 mit Beiträgen in mehreren NUM-Projektgruppen; Artikel im DZG-Magazin Synergie (Oktober 2024); Poster beim DZG-Tag in Heidelberg.
6. 2024: Zwischenevaluation in Gießen (Oktober).
7. 2025: Präsenztreffen in München (Juni) zur Planung der Weiterführung der Arbeit der AG Forschungs-IT im Folge-Vorhaben (ab 2026 unter Koordination des IDMKD Frankfurt im Rahmen von TI 12.202).

**Eigenanteil AG Kohorten:** Abstimmung mit der DZG-AG Öffentlichkeitsarbeit und der Webagentur zur Verortung des DZG-Studienregisters auf der DZG-Website; DZIF-seitige Gremienarbeit; Sprecherrolle in der AG Forschungs-IT (Vehreschild, 2022).

## b) Milestones and Deliverables

### Milestones

#### Universität zu Köln

| No. | Title | Work package | Status | Date according to the project description | Corrected date | Cor |
|---|---|---|---|---|---|---|
| 1 | Vollständige Sichtung aller in den DZG etablierten Tools und Implementierungsplan für alle AP | 1–6 | Completed | 28.02.2021 | | None |
| 2 | Fertigstellung aller Liefergegenstände für 2021 | 1–6 | Not achievable | 31.12.2021 | | Mehrere aus 2021 aufgrund Abhängig (Datensc UKK, fehl Stellungr abgeschl institutio Veränder Kohorten Frankfurt siehe Lie Tabelle. |
| 3 | Webseite DZG-übergreifendes Projektregister online | 1 | Completed | 31.12.2023 | 30.06.2024 | Prototyp studies.d |
| 4 | Datentransferdienst einsatzfähig | 2 | Completed | 30.03.2023 | 31.12.2025 | Allgemei Sharing S abgeschl produktiv Köln-Inst bereitges Kohorten Frankfurt umgezog am IDMK im Folge- 12.202. |
| 5 | Data Science Orchestrator MVP einsatzbereit | 3 | Completed | 30.06.2024 | | Prototyp Publikati DKTK. |

| No. | Title | Work package | Status | Date according to the project description | Corrected date | Cor |
|---|---|---|---|---|---|---|
| 6 | Anmeldung für zentralen Authentifizierungsdienst vorhanden | 4 | Completed | 31.03.2023 | | Föderiert DZIF↔DZ |
| 7 | DZG-übergreifender Pseudonymisierungsdienst online | 5 | Not achievable | 31.12.2023 | | Kein kon Anwendu Kontext i Machbar belegt. Schnittst in NUM-S eingebra produktiv aufgebau |

## Deliverables

### Universität zu Köln

| No. | Title | Work package | Status | Date according to the project description | Corrected date | Con |
|---|---|---|---|---|---|---|
| 1 | Evaluationsbericht Projektregister | 1 | Completed | 31.03.2021 | | None |
| 2 | Fortschrittsbericht Projektregister | 1 | Completed | 31.03.2023 | | Basisdate Portal ver 10.21961 Studienm produktiv studies.d |
| 3 | Implementierungsbericht Projektregister | 2 | Completed | 30.09.2021 | | Der DZIF-Liefergeg durch die Datensch Sicherhei UKK-Stab übergeor der TMF-S erfolgte i DZD. |

| No. | Title | Work package | Status | Date according to the project description | Corrected date | Com |
|---|---|---|---|---|---|---|
| 4 | Prozess Datentransfer mit Stellungnahme TMF AG Datenschutz | 2 | Completed | 30.09.2021 | | Der DZIF-Liefergeg durch die Datensch Sicherhei UKK-Stab übergeor der TMF-S erfolgte ir DZD. |
| 5 | Implementierungsbericht Datentransferplattform | 2 | Completed | 30.09.2023 | 31.12.2024 | Allgemeir Sharing S abgeschl dokumen Instanz ni Neuaufba Frankfurt Vorhaben |
| 6 | Prototyp eines Data Science Orchestrators, mind. 1 Analysetool | 3 | Completed | 30.09.2021 | 31.12.2024 | Prototyp Lead DKT |
| 7 | Fortschrittsbericht Data Science Orchestrator | 3 | Completed | 30.09.2023 | | None |
| 8 | Implementierungsbericht Data Science Orchestrator | 3 | Completed | 31.12.2025 | | DSO-Prot Architekt Kompone Publikatic |
| 9 | Konzept für Authentifizierungsstruktur | 4 | Completed | 30.06.2021 | | None |
| 10 | Implementierungsbericht Authentifizierungsstruktur | 5 | Completed | 31.10.2023 | 31.12.2024 | Schnittste |
| 11 | Handlungsempfehlung Record Linkage | 5 | Completed | 31.10.2021 | 31.12.2024 | Empfehlu produktiv übergreif mangels Schnittste in NUM-St eingebrac |
| 12 | Fortschrittsbericht Record Linkage | 5 | Completed | | | None |

| No. | Title | Work package | Status | Date according to the project description | Corrected date | Con |
|---|---|---|---|---|---|---|
| 13 | Implementierungsbericht Record Linkage | 5 | Completed | 31.12.2025 | | None |
| 14 | Bericht Gremienaktivität der Projektmitglieder | 6 | Completed | 31.12.2021 | | None |
| 15 | Bericht Gremienaktivität der Projektmitglieder | 6 | Completed | 31.12.2022 | | None |
| 16 | Bericht Gremienaktivität der Projektmitglieder | 6 | Completed | 31.12.2023 | | None |
| 17 | Bericht Gremienaktivität der Projektmitglieder | 6 | Completed | 31.12.2024 | | None |
| 18 | Bericht Gremienaktivität der Projektmitglieder | 6 | Completed | 31.12.2025 | | None |

**c) Most important items of the financial report**

**Personnel cost**

**Universität zu Köln**

Beschäftigungen über die Vorhabenlaufzeit (Funktionsbeschreibungen gemäß Datensparsamkeitsempfehlung):

1. Wissenschaftlicher Mitarbeiter, TV-L E13 Stufe 4–5, 100 % (anfänglich 60,5 %), 01/2021 – 12/2023
2. Wissenschaftliches Personal 1, TV-L E13 Stufe 5, 100 %, 7 Monate in 2024
3. Wissenschaftliches Personal 2, TV-L E13 Stufe 1, 100 %, 6,5 Monate in 2024
4. Wissenschaftliche Mitarbeiterin 1, TV-L E13 Stufe 5, 10,38 %, 12 Monate in 2025
5. Wissenschaftlicher Mitarbeiter 2, TV-L E13 Stufe 3, 40 %, 6 Monate in 2025
6. Wissenschaftliche Mitarbeiterin 3, TV-L E13 Stufe 2/3, 10,38 %, 3 + 9 Monate in 2025
7. Wissenschaftliche Mitarbeiterin 4, TV-L E13 Stufe 3, 10,38 %, 12 Monate in 2025
8. Wissenschaftliche Mitarbeiterin 5, TV-L E13 Stufe 1, 100 %, 12 Monate in 2025
9. Beschäftigungen in 01–02/2026 gemäß laufender Belege (Abschlussphase)

Strukturhinweis: 2021–2023 als durchgehende Vollzeitstelle für eine Person; ab 2024 Mehrpersonen-Konstellation mit teils anteiligen Stellen zur Abdeckung der fachlichen Breite (Frontend-Entwicklung, Konzeption, Datenschutz-/IT-Sicherheits-Begleitung, Gremienarbeit).

**Consumables (also contracts)**

**Universität zu Köln**

1. Softwarelizenzen
2. Server- und Cloud-Hosting (u.a. Digital Ocean)
3. Domainreservierungen
4. Auftrag an externen IT-Dienstleister (Ghostroute Consulting) für Datenversionierungs- und

Reproduzierbarkeits-Komponenten des Data Science Orchestrator
5. Auftrag an externen IT-Dienstleister im Bereich DevOps (laut Umwidmung für AP3) zur Umsetzung des Data Science Orchestrator

## Investment funds

#### Universität zu Köln

Keine.

## Travel expenses

#### Universität zu Köln

Projekttreffen mit Verbundpartnern (u.a. Heidelberg beim DKFZ 2022, Heidelberg 2023)
DZG-übergreifende AG-Treffen (u.a. Gießen Zwischenevaluation Oktober 2024, München Juni 2025)
DZIF-Jahrestagungen
Wissenschaftliche Symposien und Workshops (GMDS/TMF-Jahrestagung 2022 Köln, NUM Convention 2024, DZG-Tag 2024 Heidelberg)

## d) Necessity and adequacy of the work done

#### Universität zu Köln

Der Mittelumfang einer E13-Stelle plus Sach- und Reisekosten pro DZG war für den im Vorhaben geleisteten Aufgabenumfang angemessen, aber knapp bemessen. Er deckte konzeptionelle Abstimmung und Prototyp-Entwicklung, stieß aber für produktive DZG-weite Infrastruktur an strukturelle Grenzen.
Die selektive Vergabe von Unteraufträgen an Freelancer (Beispiel: Datenversionierungs-Komponenten des Data Science Orchestrator, mit Co-Autorenschaft in der MIE-2025-Publikation belegt) hat fokussierte Entwicklungsaufgaben ohne Aufbau von Dauerstellen adressiert.
Eine Finanzierung aus Haushaltsmitteln oder anderen Drittmitteln war nicht möglich: DZG-übergreifende Forschungs-IT-Infrastruktur ist strukturell weder durch BMBF-/DFG-Projektförderung noch durch universitäre Haushalte abzudecken; das Ziel wäre ohne DZG-spezifische Förderung nicht erreicht worden.

## e) Prospective benefit of the project, in particular the usability of the results according to the updated exploitation plan

#### Universität zu Köln

**AP1 Studienmetadatenkatalog und Basisdatensatz:** Die Plattform (studies.dzg.digital) wird über die Projektlaufzeit hinaus betrieben. Das DZPG nutzt das System für sein eigenes Studienregister nach. Der Basisdatensatz ist im MDM-Portal öffentlich zugänglich und seit 2024 verpflichtende Erhebungsgrundlage in DZG-finanzierten Studien. Erweiterungen für Patient:inneninformationen sind in der DZG-AG Patient:innenbeteiligung in Diskussion. Im Folge-Vorhaben TI 12.202 (AP2) wird die ETL-Anbindung des DZIF an das produktive Studienregister umgesetzt.
**AP2 Datentransferdienst:** Der dokumentierte, Nextcloud-basierte DZG Data Sharing Service ist nachnutzbar. Die ursprünglich in Köln geplante DZIF-Instanz wird durch den Umzug der zuständigen AG Kohorten an das IDMKD Frankfurt dort neu aufgebaut. Im Folge-Vorhaben TI 12.202 (AP4) wird auf dieser Basis eine Data Sharing Platform als übertragbare Referenzimplementierung entwickelt.
**AP3 Data Science Orchestrator:** Der Prototyp ist Grundlage für die Weiterentwicklung zu einem DZG Trusted Research Environment im Kontext von GDNG und EHDS. Im Folge-Vorhaben TI 12.202 (AP5) erfolgt die Weiterentwicklung zu einer produktiven, gesicherten Analyseumgebung mit Security Console und kontinuierlicher Pipeline für reproduzierbare Multi-DZG-Analysen.
**AP4 Authentifizierungsstruktur:** Die produktive Infrastruktur ist verfügbar und wird DZG-weit für neue Dienste genutzt. Im Folge-Vorhaben TI 12.202 (AP3) wird ein Integrationskonzept für NUM Hub

und IAM4NFDI entwickelt.

**AP5 Record Linkage:** Die im Projekt gewonnenen Schnittstellenanforderungen wurden in die NUM-Strukturen eingebracht (DZG-AG-Jahresbericht 2024). Die durchgeführten SMPC-Experimente und der dokumentierte Testdatensatz sind für künftige Pilotierungen nachnutzbar, sollten sich entsprechende Anwendungsfälle ergeben.

**AP6 Außendarstellung:** Sichtbarkeit der DZG im Bereich Forschungs-IT ist durch GMDS-Workshop, NUM-Convention-Beiträge, Synergie-Artikel und DZG-Tag-Poster erhöht. Die Außenvertretungstätigkeit wird im Folge-Vorhaben fortgesetzt.

Sämtliche im Vorhaben entwickelte Software wurde unter Open-Source-Lizenzen veröffentlicht, wodurch eine breite Nachnutzung durch andere Forschungsinfrastrukturen, universitätsmedizinische Einrichtungen, die NFDI und potenziell auch durch Unternehmen ohne Lizenzbarrieren ermöglicht wird. Die unmittelbare Anschlussverwertung im DZIF erfolgt neben TI 12.202 auch über die TI BBD (Aufbau der DZIF-IT-Plattform).

### f) Progress/Advances in the field of the project by third parties during implementation of the project

#### Universität zu Köln

Während der Vorhabenlaufzeit haben sich mehrere parallele Initiativen entwickelt, die für das Vorhaben relevant sind und in dessen Arbeit eingeflossen sind:

1. **Netzwerk Universitätsmedizin (NUM):** Aufbau zentraler Infrastrukturen (NUM Hub, NUM-AAI, ID- und Consent-Management). Die DZG-AG Forschungs-IT hat Anforderungen aus AP4 und AP5 in die NUM-Projektgruppen eingebracht.
2. **NFDI4Health:** Aufbau des Study Hub; Kooperationsprojekt mit der AG zur bidirektionalen Anbindung des DZG-Studienmetadatenkatalogs.
3. **IAM4NFDI:** Entwicklung nationaler Authentifizierungsstandards; Abstimmung erforderlich für die Skalierung der DZG-Lösung.
4. **Medizininformatik-Initiative (MII):** Entwicklung von Kerndatensätzen, die als Referenzen für den DZG-Basisdatensatz genutzt wurden.
5. **Gesundheitsdatennutzungsgesetz (GDNG) und EHDS:** Neue regulatorische Anforderungen, die die Weiterentwicklung des DSO zu einem TRE motivieren.

Keine dieser Entwicklungen hat die ursprüngliche Zielsetzung des Vorhabens obsolet gemacht; vielmehr haben sie die Notwendigkeit der DZG-übergreifenden Infrastruktur bestätigt und konkrete Schnittstellenanforderungen geliefert, die im Folge-Vorhaben adressiert werden.

### g) Successful or planned publication of the results according to No. 5 of the “Nebenbestimmungen für Zuwendungen” (NABF/NKBF 2017)

#### Universität zu Köln

DZG-Basisdatensatz im MDM-Portal: medical-data-models.org/45851 (DOI 10.21961/mdm:45851, veröffentlicht am 19.09.2023, Lizenz CC BY-NC-SA 4.0)

DZG Studienübersicht: studies.dzg.digital

DZG Data Sharing Service: git.apps.dzd-ev.org/dzdtools/dzddatasharingservice

**Kern J, Brechtel M, Kaluza P, Kussel T, Vehreschild JJ, Lablans M.** *Data Science Orchestrator: A Containerized Trusted Research Environment for Flexible and Secure Analytical Pipelines.* Studies in Health Technology and Informatics, Vol. 327 (Intelligent Health Systems – From Technology to Data and Knowledge), Andrikopoulou E et al. (Eds.), IOS Press 2025, S. 698–702. DOI: 10.3233/SHTI250438. Open Access (CC BY-NC 4.0).

GMDS/TMF 2022, Köln: Workshop "Research IT Collaboration and Challenges across German Centres for Health Research (DZG)"

DZG-Magazin Synergie, Oktober 2024: Artikel zur DZG-Forschungs-IT

DZG-Tag 2024 Heidelberg: Poster zu Arbeitsergebnissen der DZG-AG Forschungs-IT