



Oliver Kohlbacher

University of Tübingen and University Hospital Tübingen

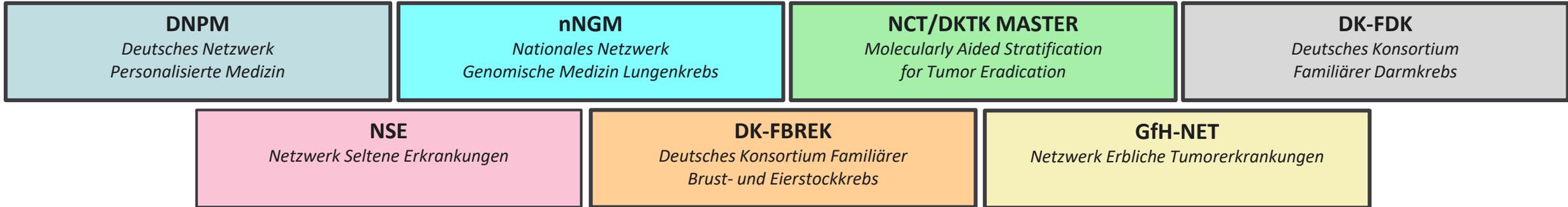
EBERHARD KARLS
**UNIVERSITÄT
TÜBINGEN**



**Universitätsklinikum
Tübingen**

Sieben Versorgungsnetzwerke

Vorreiter in der genomischen Medizin

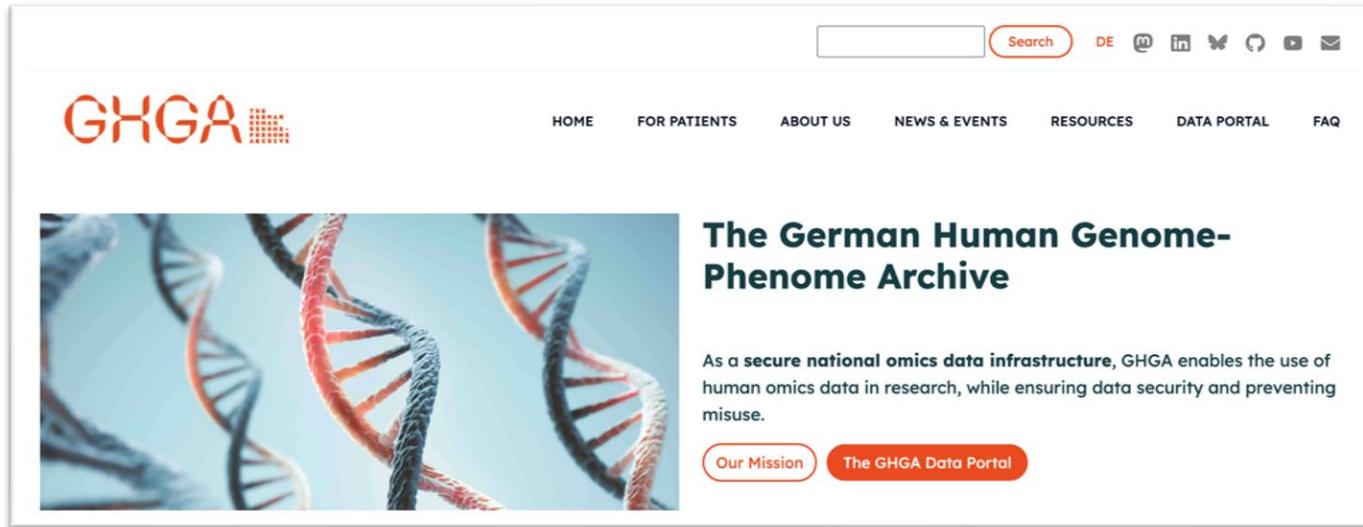


- Sieben Versorgungsnetzwerke setzen in kleineren Verbänden bereits seit einiger Zeit genomische Medizin an der Schnittstelle zur Forschung um
- Finanzierung aus verschiedenen Quellen – in der Regel Forschungsmittel
- Netzwerke haben jeweils eigene
 - > Einschlusskriterien
 - > Zu dokumentierende Datensätze
 - > Dateninfrastrukturen
 - > Teilnehmende Einrichtungen
- **Idee**
 - **genomDE nutzt existierende Strukturen der Netzwerke (technisch und organisatorisch)**
 - **Netzwerke werden für alle Universitätsklinika mit Zulassung zum genomDE MV geöffnet**



German Human Genome-Phenome Archive (GHGA)

Nationale Forschungsdateninfrastruktur für Genomdaten



In Zusammenarbeit mit



Gefördert durch



- Das Deutsche Humangenom-Phänomarchiv ist eine Nationale Forschungsdateninfrastruktur (NFDI)
- GHGA besitzt sechs Datenhubs – Rechenzentren, die im großen Maßstab Genomdaten zur Forschung verwalten
- **Idee**
 - **Data Hubs in GHGA besitzen die Infrastruktur und Personal zum Management von Genomdaten**
 - **Nutzung dieser Infrastruktur schafft Synergien und senkt Kosten für genomDE**



genomDE - Infrastrukturkomponenten

Klinischer Datenknoten (KDK)

Speichert **strukturierte klein. Daten**

ermöglicht die Qualitätskontrolle und Datenbereitstellung für klinische Daten über existierende Netzwerke

Genomrechenzentrum (GRZ)

speichert **Genomdaten**

ermöglicht Analyse und Bereitstellung der Rohdaten über Forschungsdateninfrastruktur

Datendienst (DD)

stellt **prozessierte Daten**
Aus KDK und GRZ bereit

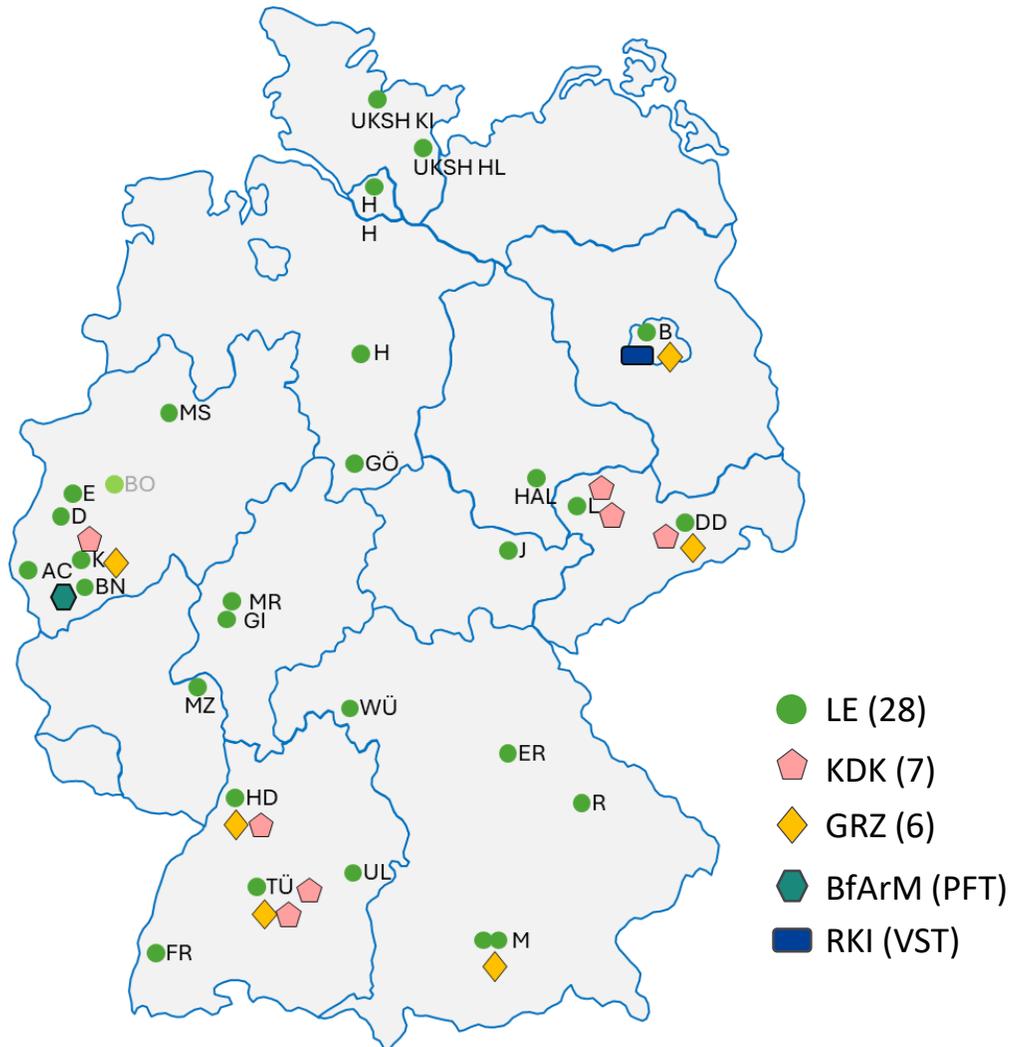
- noch in Vorbereitung-

- **Vorteil:** Betriebsstart durch Nutzung bereits existierender Strukturen etwa 2 – 3 Jahre früher und mit geringeren Kosten möglich
- **Nachteil:** Harmonisierung und Konvergenz der Infrastruktur aufwändig und Abstimmung sehr komplex
- **Abwägung:** Früherer Zugang zum Modellvorhaben für Patient:innen rechtfertigen Komplexität und geringeren Harmonisierungsstand

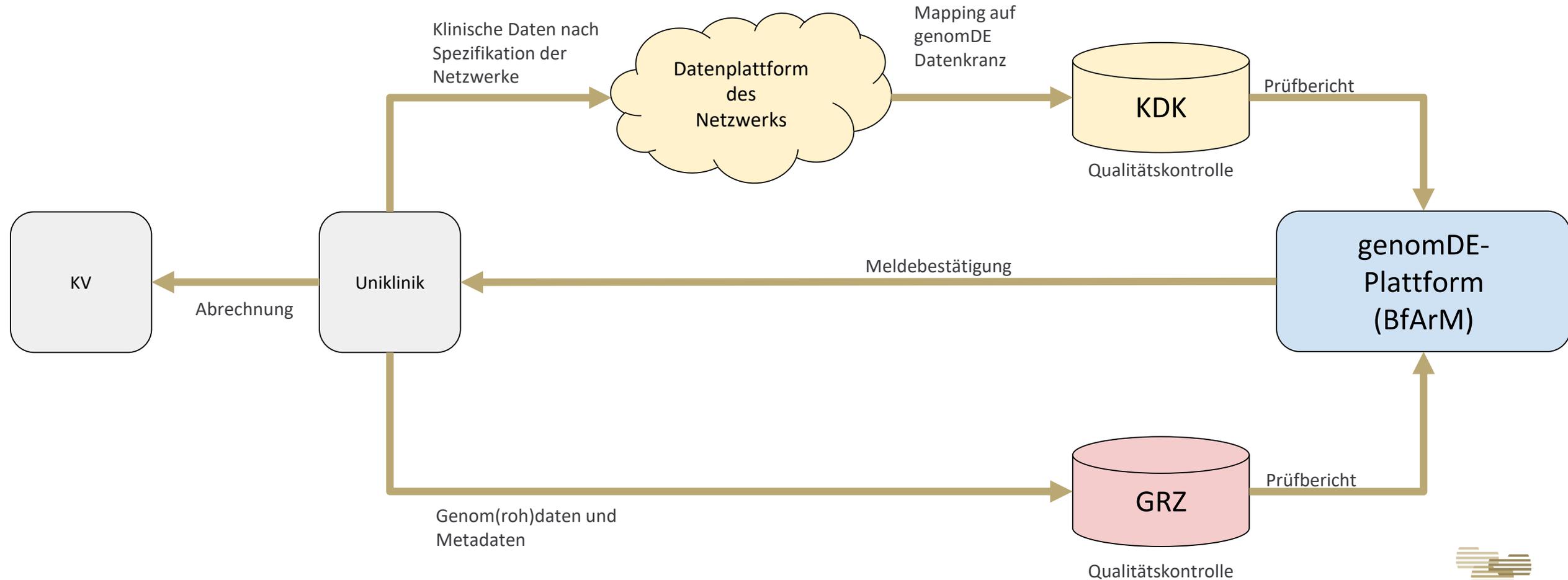


Zulassung GRZ und KDK

- Infrastruktur setzt auf existierende Komponenten – mit nachgewiesener Eignung
- Ausschreibungsverfahren: Leistungserbringer bewarben sich um Zulassung als KDK und GRZ
- Zugelassen vom BfArM
 - Sieben KDK (je einer pro Netzwerk)
 - Sechs GRZ (GHGA Data Hubs)
- Verteilte Dateninfrastruktur vom Plattformträger mit der Umsetzung des genomDE MV beauftragt
- Betreiber haben Dienstleistungsverträge mit BfArM geschlossen



Datenflüsse Modellvorhaben



Modellvorhaben – Die Daten

- §64e SGB V spezifiziert die Datenkategorien (Person, Genom, klinisch, ...)
- Genomdatenverordnung (GenDV) definiert einzelne Datenitems
- Technische Spezifikation (JSON-Schema) des BfArM definiert technische Umsetzung

§64e SGB V

(10) Die Leistungserbringer übermitteln für die in Absatz 11 Satz 3 genannten Zwecke für jeden teilnehmenden Versicherten

- 1.an die Vertrauensstelle die von ihnen erstellte Arbeitsnummer und die Krankenversichertennummer im Sinne des § 290,
- 2.an ein Genomrechenzentrum die Vorgangsnummer und die Daten der Genomsequenzierung,
- 3.an einen klinischen Datenknoten die Vorgangsnummer, die klinischen Daten und die Daten der jeweiligen Einwilligung.

Die klinischen Daten umfassen die Angaben zu Alter, Geschlecht und Kreisschlüssel, die Daten der Phänotypisierung sowie die Daten zum Behandlungsverlauf.

Für eine in Absatz 9c Satz 7 Nummer 1 genannte Fallidentifizierung übermittelt

- 1.der ersuchende Leistungserbringer an die Vertrauensstelle das über einen Datendienst bereitgestellte Genomdatensatz- oder Klinikdatensatzpseudonym,
- 2.der ersuchte Leistungserbringer an die Vertrauensstelle die nach Absatz 9c Satz 3 Nummer 4 generierte zufällige Vorgangsnummer, die ihm vom ersuchenden Leistungserbringer mitgeteilt wurde.

[...]

GenDV

- 13 - Drucksache 257/24

- bb) Monat und Jahr des ersten Kontakts mit einem spezialisierten Zentrum für Seltene Erkrankungen
- VII. Klinische Daten zur Nachverfolgung (zur Übermittlung an einen klinischen Datenknoten)
1. für onkologische Erkrankungen
 - a) Umsetzung einer Therapieempfehlung der Fallkonferenz (einschließlich der präventiven Maßnahmen bei erblichen Tumorerkrankungen)
 - b) Therapie
 - aa) Datum Start der Therapie
 - bb) Datum Ende der Therapie
 - cc) Art der Therapie
 - dd) Wirkstoff
 - c) Therapieansprechen (best response)
 - aa) Erhebungsdatum
 - bb) Datenquelle Therapieansprechen
 - cc) Progression
 - dd) Stabile Erkrankung
 - ee) Partielle Remission
 - ff) Komplette Remission
 - gg) ergänzend für erbliche Tumorerkrankungen: Auftreten meta-chroner Tumorerkrankungen mit Diagnoseschlüssel
 - d) ECOG Performance Status
 - e) Allgemeiner Status zur klinischen Nachverfolgung
 - aa) Vitalstatus
 - bb) Datum des letzten Kontakts
 - cc) Sterbetag, Sterbemonat und Sterbejahr
 2. für seltene Erkrankungen
 - a) Dokumentation Wiedervorstellung
 - aa) Datum
 - bb) Angaben zu Änderungen im Phänotyp auf der Grundlage der HPO

„Datenkranz“

lanC	Vorgangsnummer	String	%a-FA-F0-9954\$	1	Ein einzigartiger 32-byte langer Code, dargestellt als Hex-String mit einer Länge von 64 Zeichen. Der lanC bezieht sich immer auf den Indexpatienten (Indexierungsebene der Fall). Für jede Meldung zum Fall ist eine neue Vorgangsnummer zu übermitteln.
localCaseId	Lokale Fallbezeichnung zur lokalen Identifikation	String		0..1	Lokale Fallbezeichnung zur lokalen Identifikation, Optionale Angabe eines lokalen Fallbezeichners (nur in Phase 0). Der localCaseId für die Übermittlung an die KDK darf nicht mit localCaseId zur Übermittlung an das GRZ identisch sein
gender	Geschlecht	Enum	{male, female, other, unknown}	1	Angabe des Geschlechts (soziales Geschlecht)
birthDate	Geburtsdatum	Datum	date	1	Geburtsdatum als Monat und Jahr, Datum (kurz) nach ISO 8601 (YYYY-MM)
addressAGS	Ämlicher Gemeindeschlüssel	String(5)	12345 Bundesland (2 Stellen), Regierungsbezirk (1 Stelle), Kreis (2 Stellen)	1	5 Stellen des Ämlichen Gemeindeschlüssels https://www.destatis.de/Files/Downloads/Statistik/IGS hier = unbekannt oder kein Wohnsitz in Deutschland
decisionInclude	Entscheidung der multidisziplinären Fallkonferenz zur molekulargenetische n Diagnostik	Bool		0..1	Entscheidung der multidisziplinären Fallkonferenz zur molekulargenetischen Diagnostik

```
"birthDate": {
  "type": "string",
  "description": "Date of birth without the day. Format YYYY-MM",
  "pattern": "^[0-9]{4}-(0[1-9]|1[0-2])$"
},
"addressAGS": {
  "type": "string",
  "description": "Official municipal code of the patient's residence."
},
```

Beispiel: Kerndatensatz Seltene Erkrankungen (NSE)

MV GenomSeq Metadata

- Transfer TAN
- Consent information
- Submission (e.g., initial, follow-up)
- ...

Phenotype

- Date of documentation
- HPO codes
- ...

Diagnostic Report

- Type of sequencing (WGS, long/short read)
- Sequencing platform
- Variant information
- ...

Follow-Up

- Follow-up date
- ...

Patient

- Gender
- DOB
- Health insurance
- ...

Diagnosis

- Onset data
- Genetic evidence level
- Diagnosis codes (ICD10, ORDO, Alpha-ID-SE)
- ...

Therapy History

- Therapy type
- Medication
- Date
- ...

Care Plan

- Therapy recommendations
- Study enrolment
- Re-evaluation recommendation
- ...

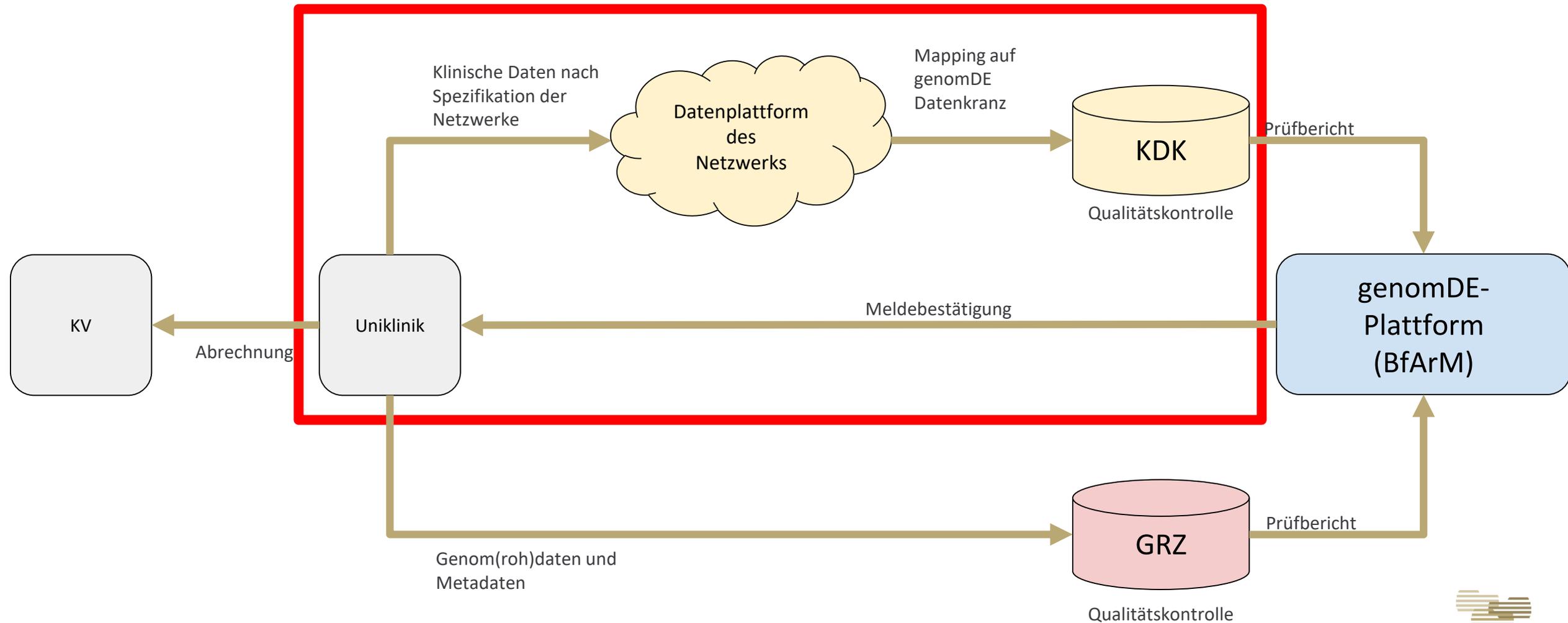
(+ a few smaller details – total around 150 data items)

Detaillierte Beschreibung:

<https://ibmi-ut.atlassian.net/wiki/spaces/DRD/pages/1474938/Data+Model+-+SE+dip>



Bereitstellung klinischer Daten



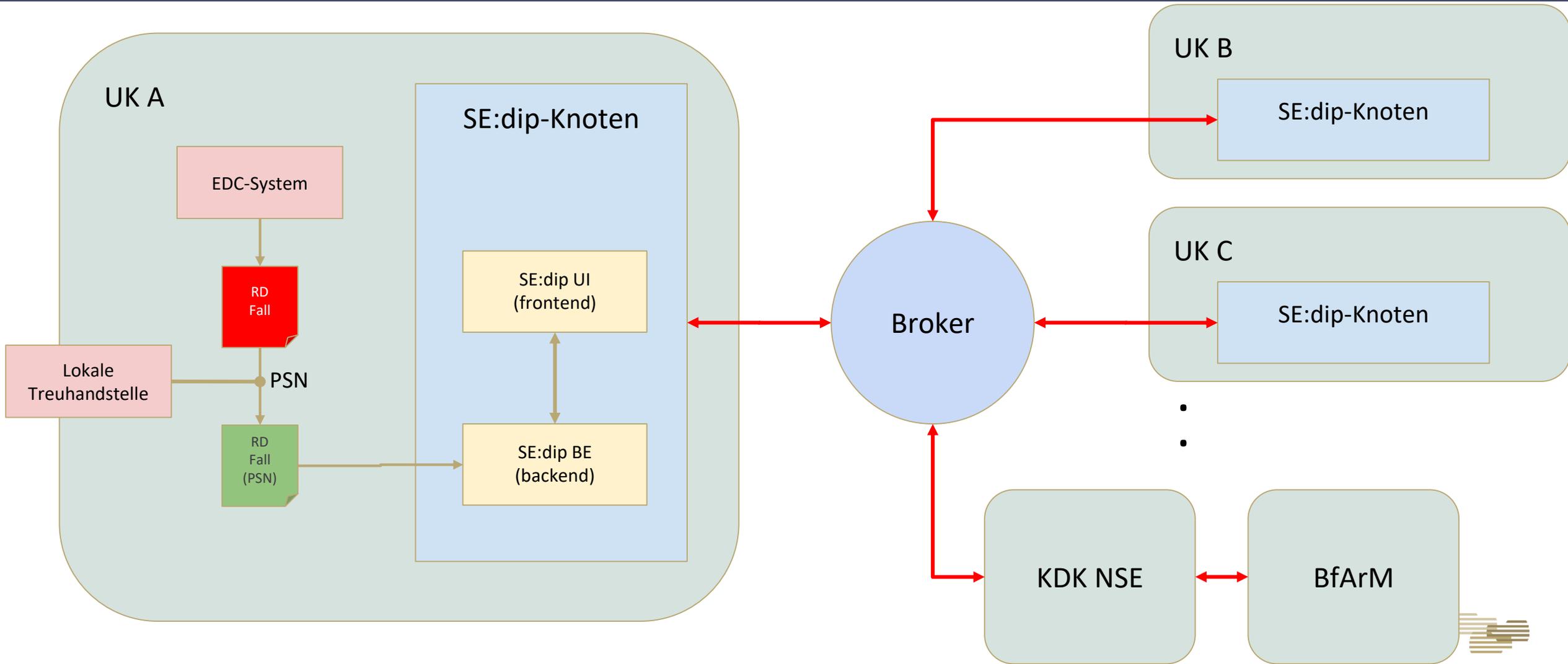
REDCap – Fallback für lokale Datenerhebung

- Integration der Datenerhebung in die klinischen Primärsysteme an vielen Standorten schwierig (und langsam)
- REDCap ist ein Electronic Data Capture (EDC)-System das viele UK nutzen
- System ist sehr einfach und schnell konfigurierbar – ideal für frühe Phasen einer Datenspezifikation
- Daten können einfach in Webformulare eingetragen werden
- Schnittstellen für Im-/Export sind einfach ergänzbar

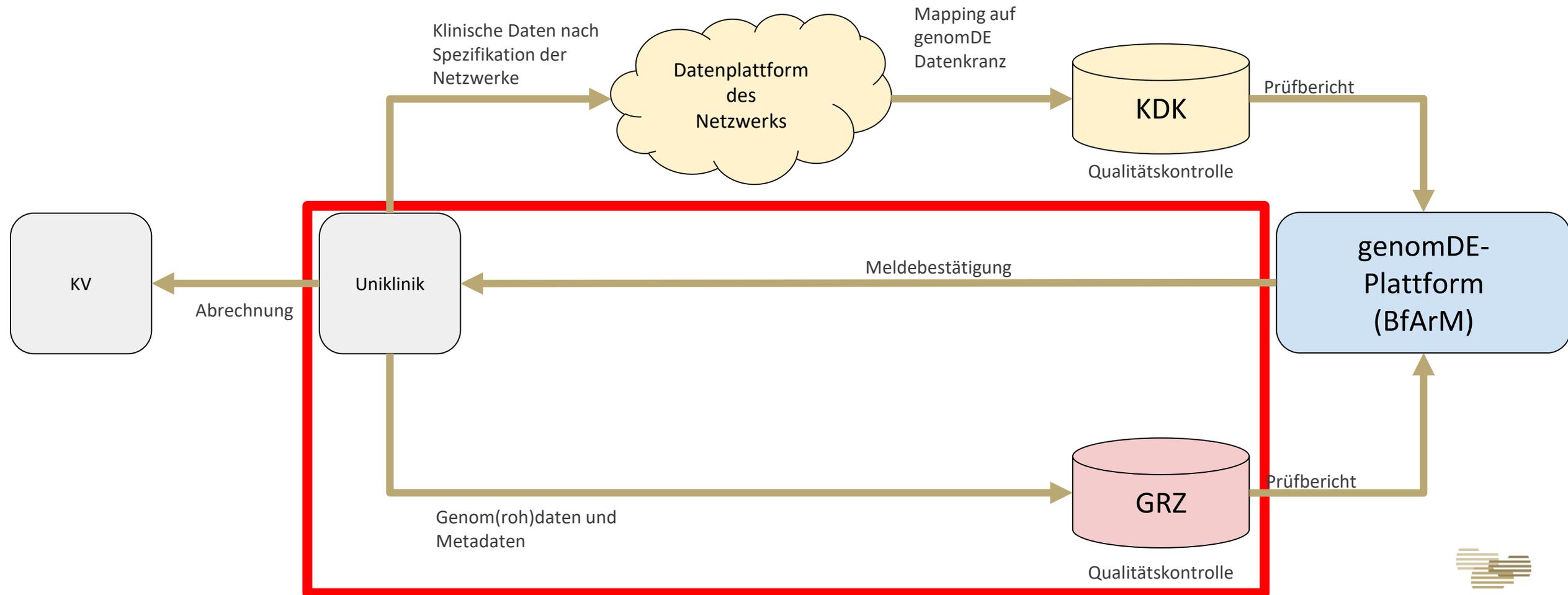
The screenshot displays the REDCap web interface. The top section, 'Datensatz Status Übersicht (Alle Datensätze)', shows a table of data records with columns for 'Pseudonym', 'III. Daten der Einwilligung und Teilnahmeerklärung', and 'IV. Allgemeine Daten zum Modellvorhaben'. Below this is a 'VII. Klinische Daten zur Nachverfolgung' form for a specific pseudonym. The form includes a table for 'Bestehenden bearbeiten Pseudonym 1' with columns for 'Datum', 'Phänotypisierung auf der Grundlage der HPO', 'Änderung des HPO', 'GMFCS', 'Diagnosestellung erfolgt', 'Ergänzende Daten zum Verlauf der Erkrankung', and 'Form Status'. The table shows entries for dates 07-10-2024, 08-10-2024, 08-10-2024, and 22-10-2024. Below the table, there are fields for 'Vitalstatus beim letzten Kontakt' and 'Form Status'.



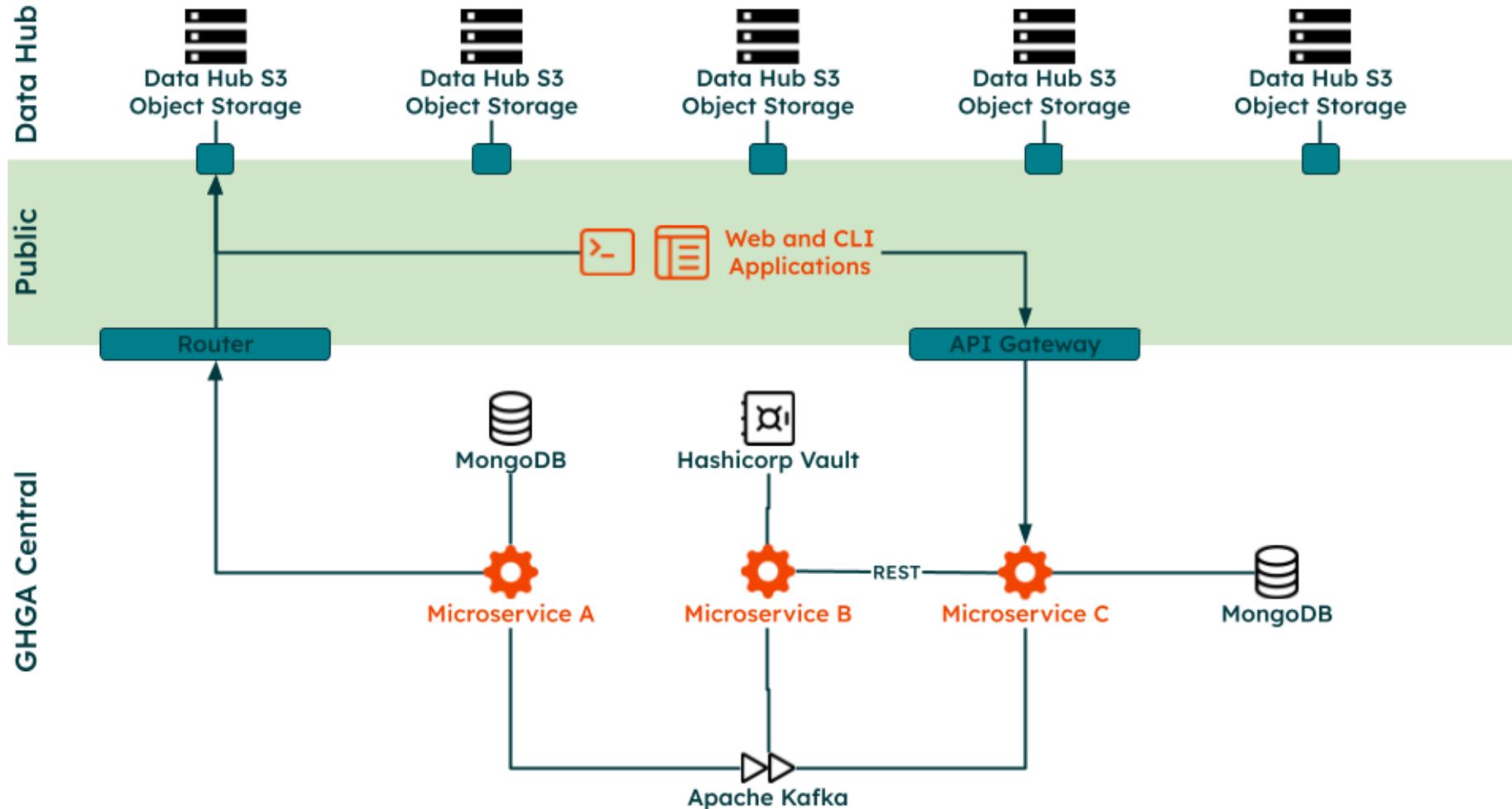
SE:dip – Architektur



Bereitstellung Genomdaten



GHGA – Technische Architektur

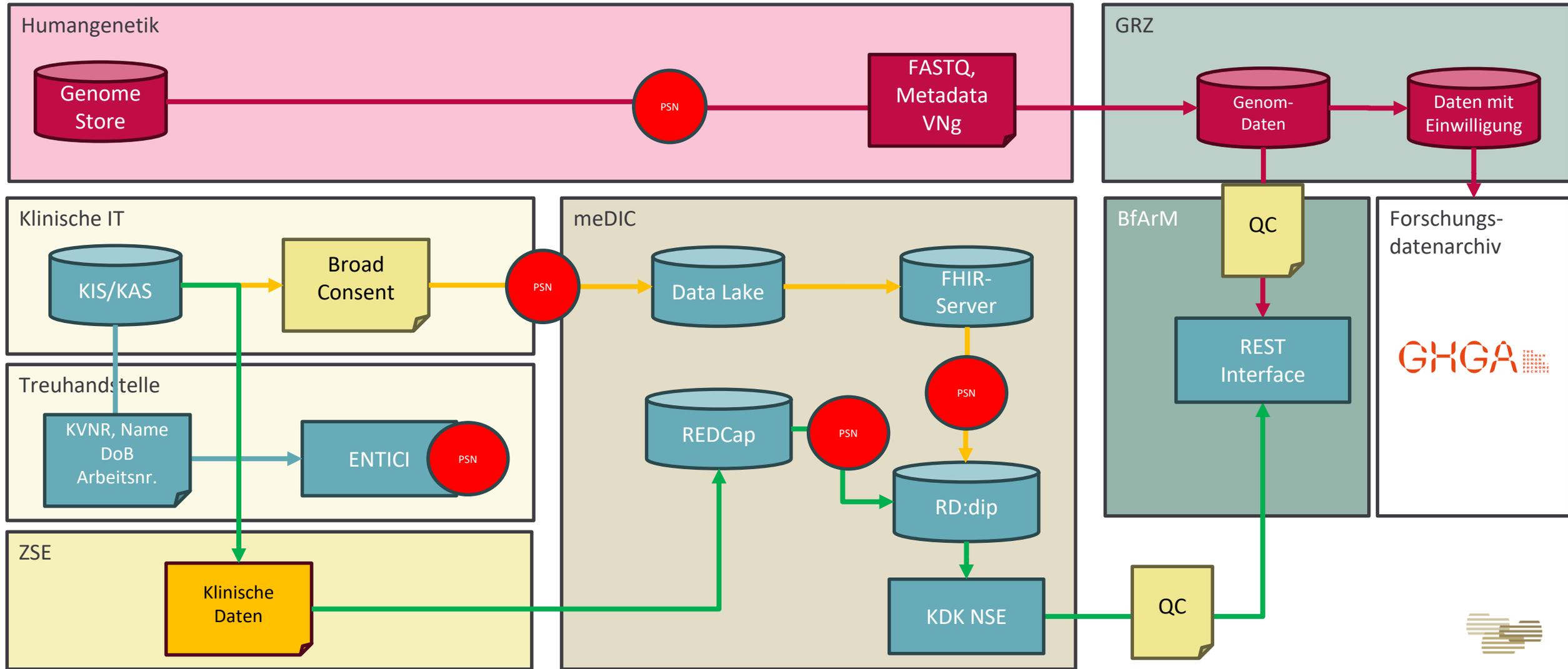


Qualitätskontrolle der Daten

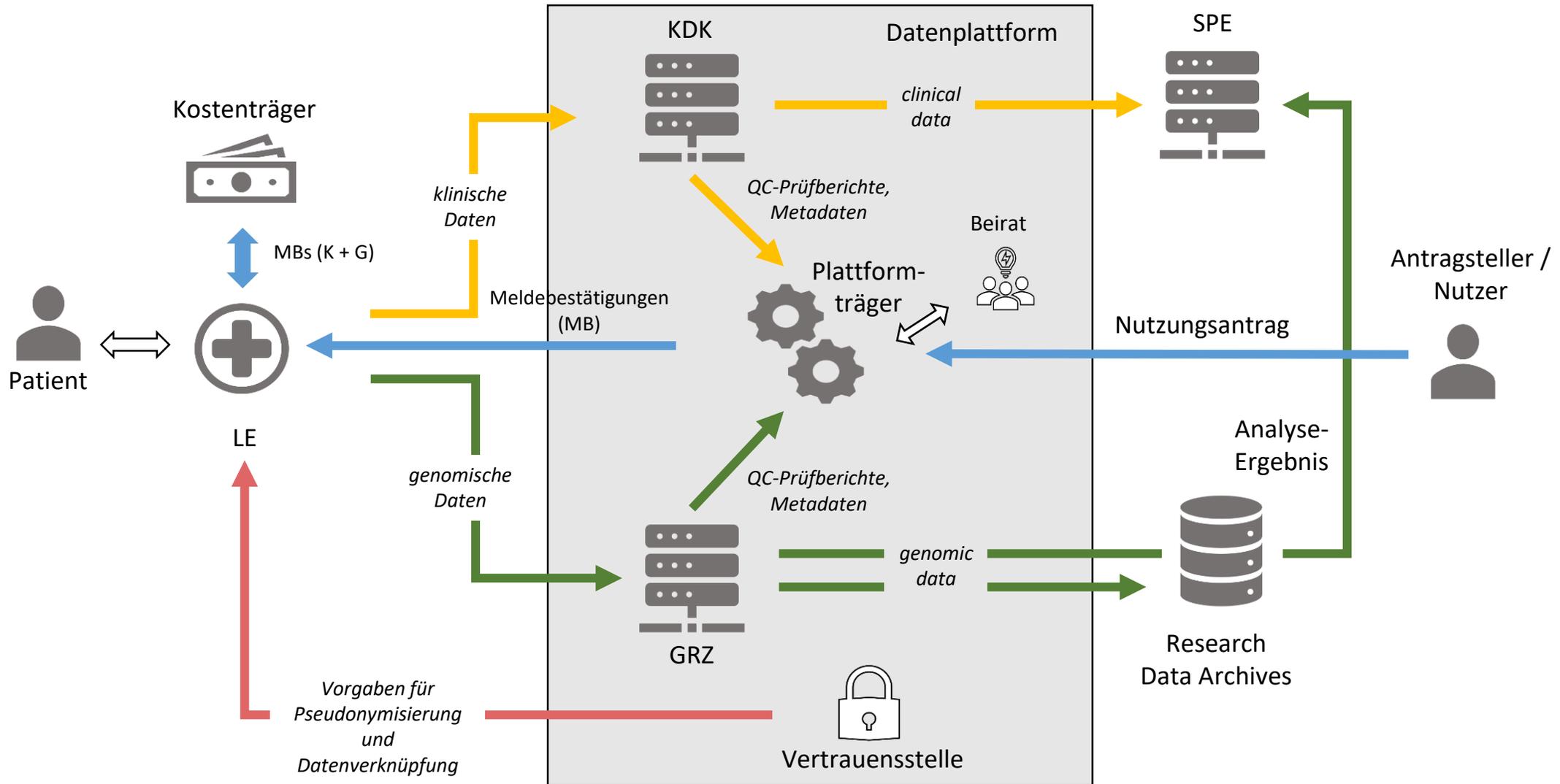
- BfArM hat **Qualitätskriterien** für Genomdaten definiert, die erfüllt sein müssen
- GDCs müssen dem BfArM für jeden Fall Qualitätskontrollinformationen im Rahmen ihrer Verarbeitungspflichten bereitstellen
- **Routinemäßige QC** bei JEDER Einreichung: Korrektheit der Übertragung, berichtete Sequenziertiefe, ...
- **Stichprobenartige vertiefte QC** (2% der Einreichungen jedes Einreichers)
 - % der Basen mit Q30 oder höher
 - Berechnete Sequenziertiefe
 - Anteil der Zielregionen mit minimaler Sequenziertiefe
- GRZ stellen eine **QC-Pipeline** zur Verfügung, die die benötigten QC-Parameter bestimmen kann
- **Hoher Rechenaufwand:** QC für Ganzgenom ca. 800 CPU-Stunden



Beispieldatenfluss - NSE in Tübingen



Vision Gesamtinfrastruktur



Nächste Schritte

- **Betriebsstart**
 - Infrastruktur startet in diesen Tagen den Produktivbetrieb
 - Technische Bereitschaft wird schrittweise erreicht werden
 - Wir erwarten die Abrechnung erster Datensätze bis zum Ende des Monats
 - Onboarding der Leistungserbringer wird sicherlich noch einige Monate dauern
- **Konsolidierung der Infrastruktur**
 - Anpassung von Prozessen
 - Harmonisierung mit MII-Kerndatensatz
 - Vereinfachung der Datenflüsse
- **Ausbau der Infrastruktur**
 - Erweiterung der sicheren Ausführungsumgebungen (SPEs)
 - Implementierung erster Datendienste



Acknowledgements

University Hospital Tübingen

Lucien Clin, Peter Placzek, Michaela Hardt, David Hieber

Nisar Malek, Yvonne Möller, Lisa Brenner

Olaf Rieß, Stephan Ossowski, Marc Sturm, Monika Glauch

GHGA/NFDI

Oliver Stegle, Jan Eufinger, Koray Kirli

Susanne Motameny, Simon Parker, Till Hartmann, Julia Leimeister, Florian Hölzlwimmer. Travis Wrightsman

The whole data hub & GDC operations team

genomDE

All working groups – but especially

Thomas Berlage, Michael Krawczak, Sebastian C. Semler



BfArM

Andreas Till, Catharina Scholl, Luca Pötschke,
Britta Hänisch

RKI

Rüdiger Dölle, Anna Lübbe

