Bundesinstitut
für Arzneimittel
und Medizinprodukte

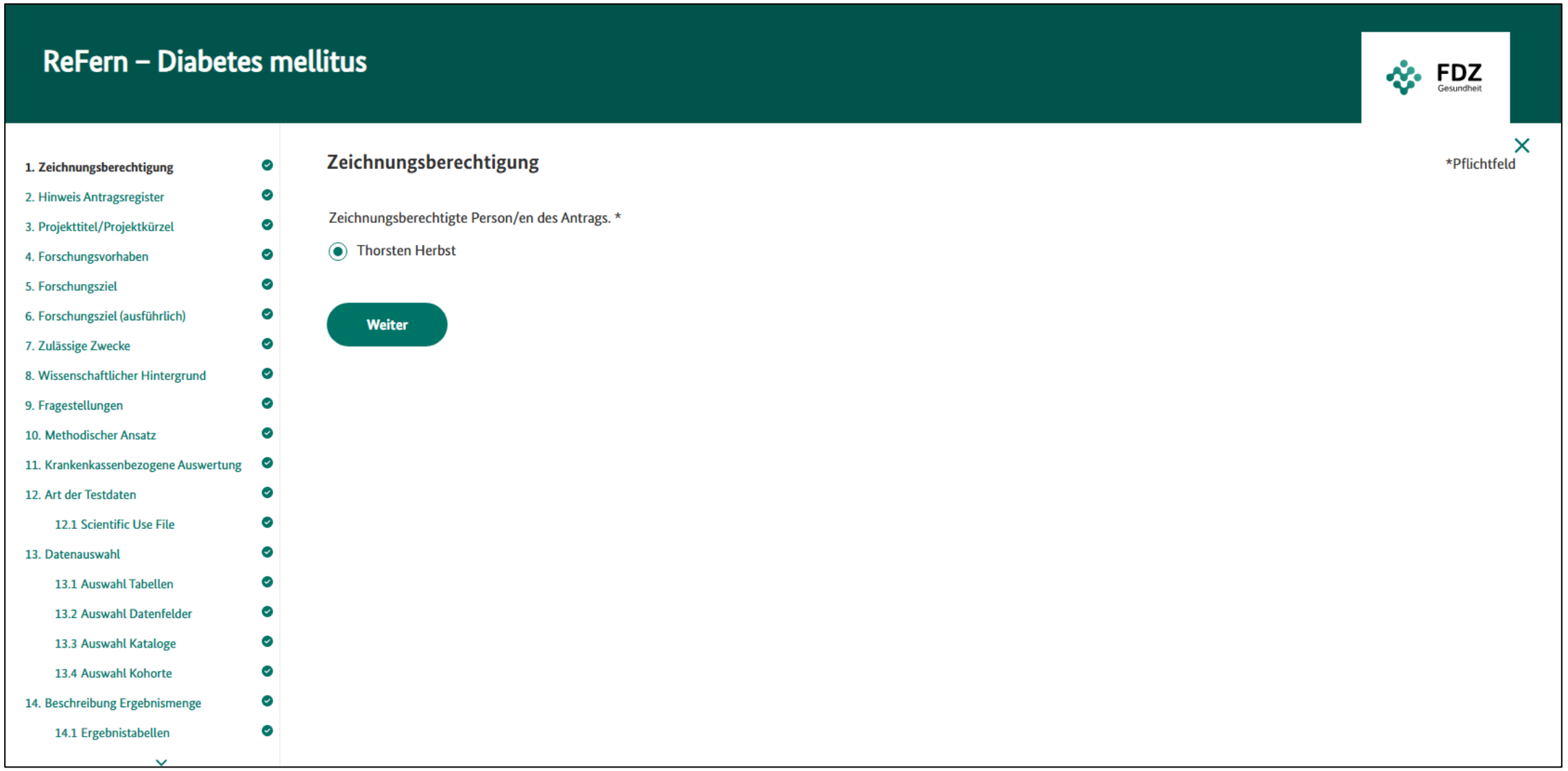
Beispiel-Antragsentwurf Forschungsdatenzentrum Gesundheit

FDZ Gesundheit Portal: Antragsentwurf am Beispiel des ReFern-Projekts

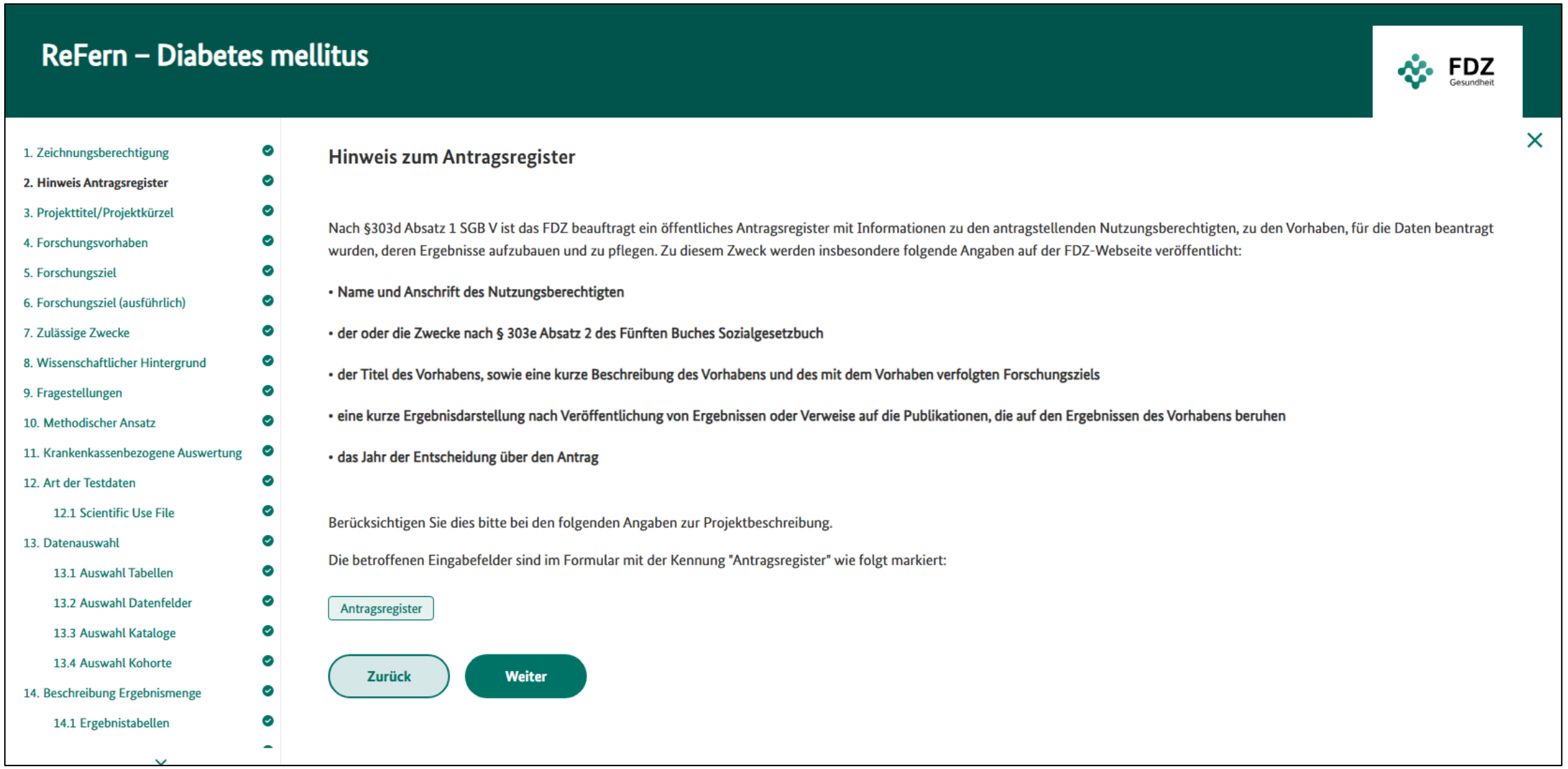
- Nutzungsberechtigte müssen ein digitales Antragsformular ausfüllen, um Zugang zu den Daten zu erhalten; dieser wird von FDZ-Mitarbeitenden geprüft und bewilligt.
- Zur Vorbereitung und Unterstützung wird hier ein **beispielhafter** Antragsentwurf vorgestellt.
- Der Entwurf orientiert sich am ReFern Projekt – einem vom Bundesministerium für Gesundheit (BMG) geförderten, gemeinsamen Vorhaben des Robert-Koch-Instituts (RKI) und des Bundesinstitut für Arzneimittel und Medizinprodukte (BfArM). Ziel des Projekts ist die Erstellung von Referenzauswertungen zu Public-Health-relevanter Erkrankungen auf Basis von Routinedaten.
    - Mehr Informationen zum Projekt finden Sie in der dazugehörigen Publikation (DOI: https://doi.org/10.1007/s00103-023-03821-1)
- Das Antragsportal ist noch im Aufbau. Die hier dargestellten Schritte können sich daher mit dem späteren Livegang noch ändern.

Bundesinstitut
für Arzneimittel
und Medizinprodukte

Schrittweiser Ablauf des Antragsentwurfs



Auswahl der **zeichnungsberechtigten Person** der Institution, die den Antrag zur inhaltlichen Prüfung freigibt oder die Freigabe ablehnt, falls der Antragsentwurf überarbeitet werden soll.



Das FDZ legt ein **öffentliches** Antragsregister an, in dem **alle** abgeschlossene Anträge gelistet werden.

ReFern – Diabetes mellitus

FDZ Gesundheit

1. Zeichnungsberechtigung
2. Hinweis Antragsregister
3. **Projekttitel/Projektkürzel**
4. Forschungsvorhaben
5. Forschungsziel
6. Forschungsziel (ausführlich)
7. Zulässige Zwecke
8. Wissenschaftlicher Hintergrund
9. Fragestellungen
10. Methodischer Ansatz
11. Krankenkassenbezogene Auswertung
12. Art der Testdaten
    - 12.1 Scientific Use File
13. Datenauswahl
    - 13.1 Auswahl Tabellen
    - 13.2 Auswahl Datenfelder
    - 13.3 Auswahl Kataloge
    - 13.4 Auswahl Kohorte
14. Beschreibung Ergebnismenge
    - 14.1 Ergebnistabellen

Projekttitel / Projektkürzel

*Pflichtfeld

Projekttitel *

Antragsregister

Referenzauswertungen für die Schätzung von Prävalenz, Inzidenz und Mortalität Public-Health-relevanter Erkrankungen auf Basis von Versorgungsdaten am Beispiel von Diabetes Mellitus in den Jahren 2009-2018

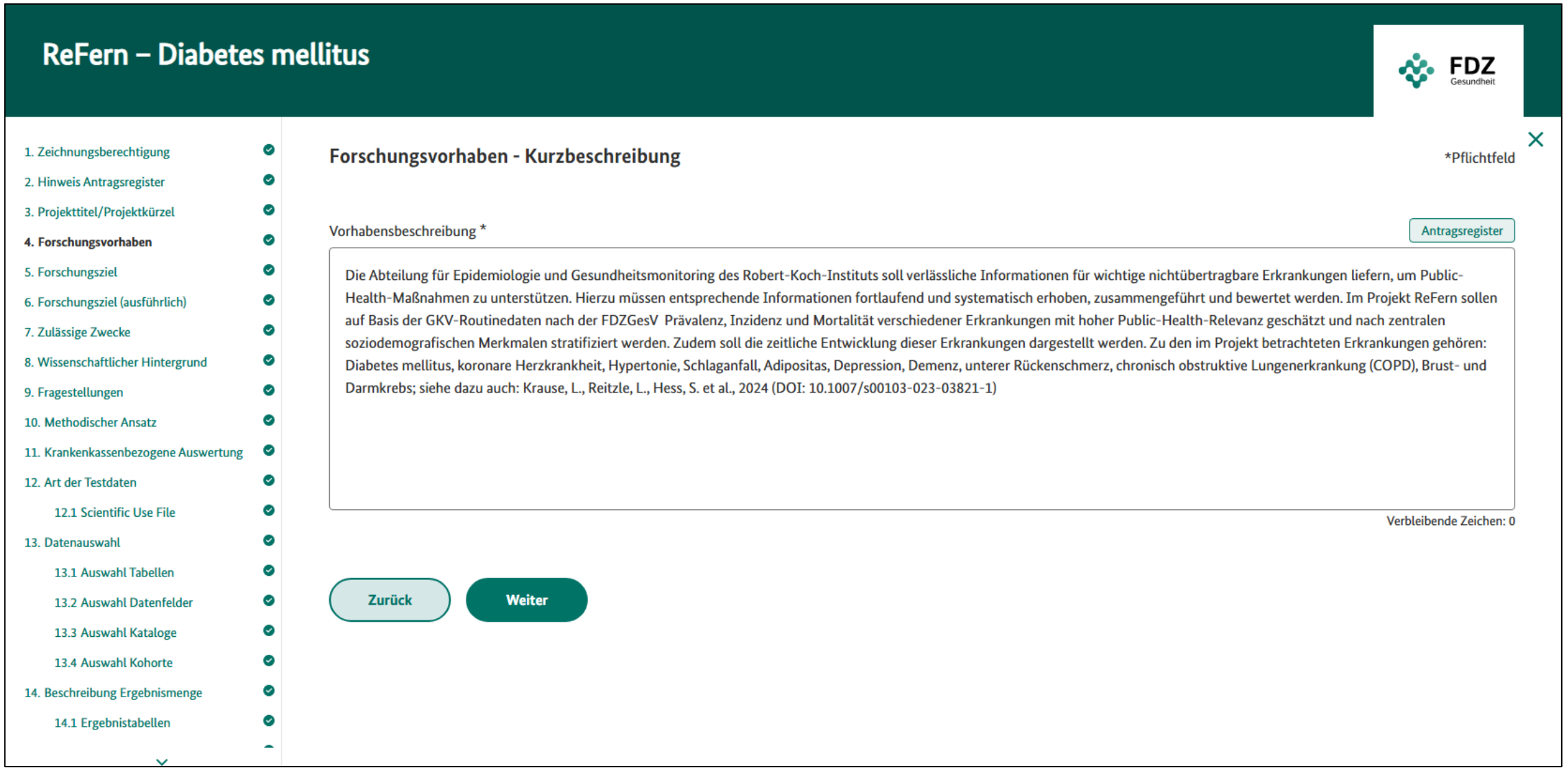
Verbleibende Zeichen: 196

Projektkürzel *

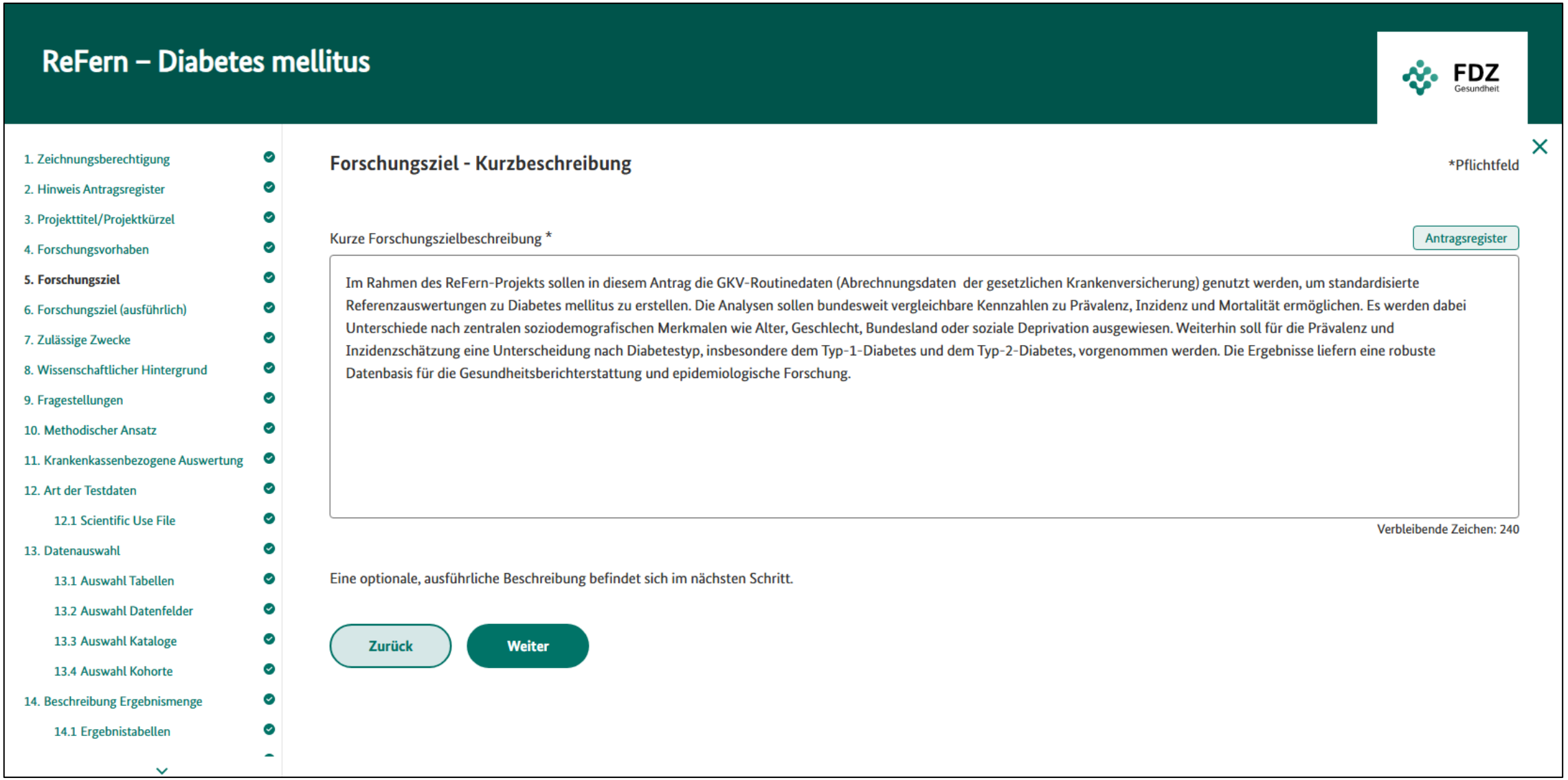
ReFern – Diabetes mellitus

Verbleibende Zeichen: 34

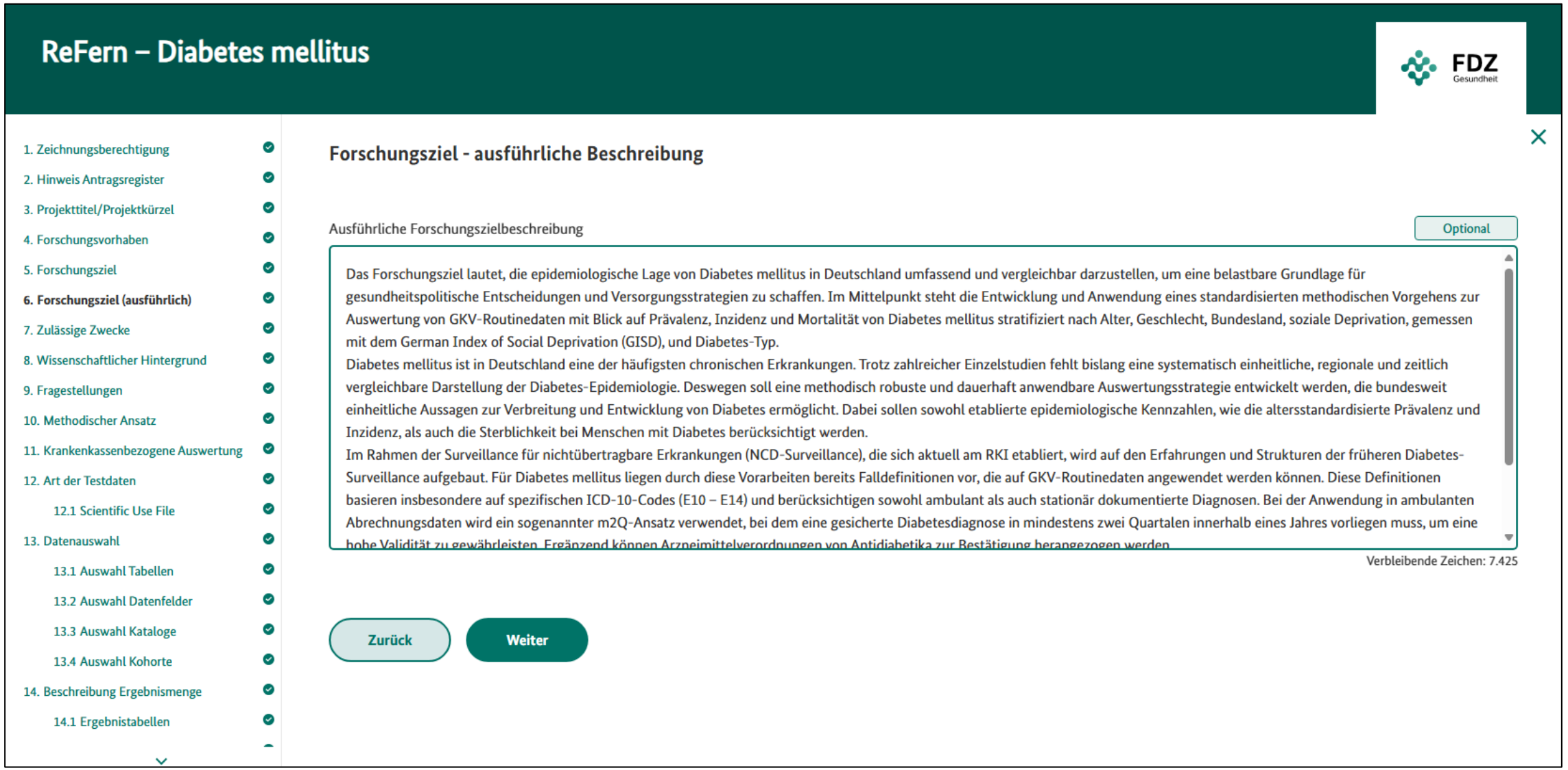
Zurück | Weiter



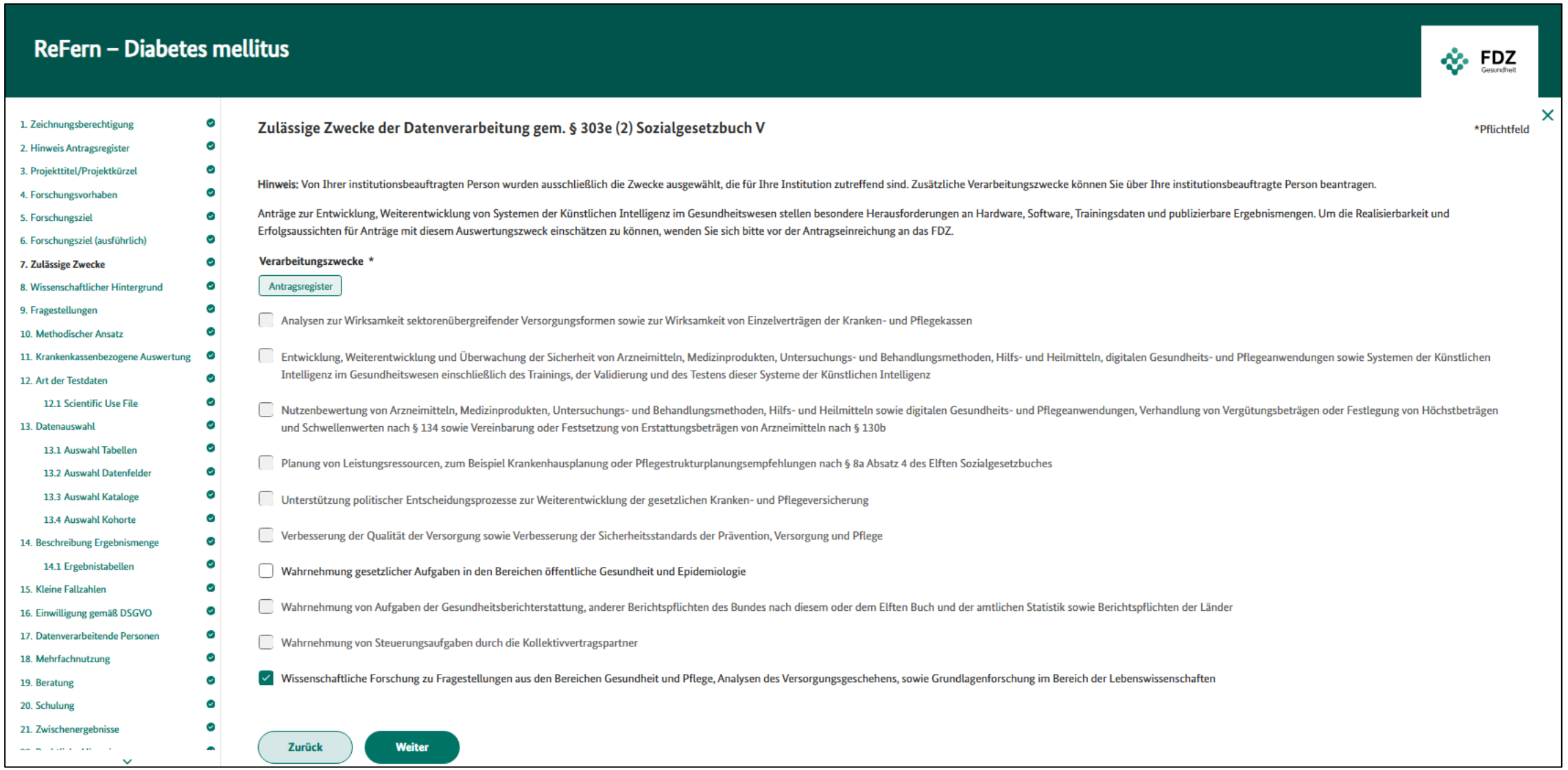
Beim Forschungsvorhaben soll **kurz** beschrieben werden, in welchem **Forschungskontext** das geplante Projekt steht, gegebenenfalls mit einer Erläuterung des thematischen Schwerpunkts der Forschungsgruppe sowie der Rolle des Projekts.



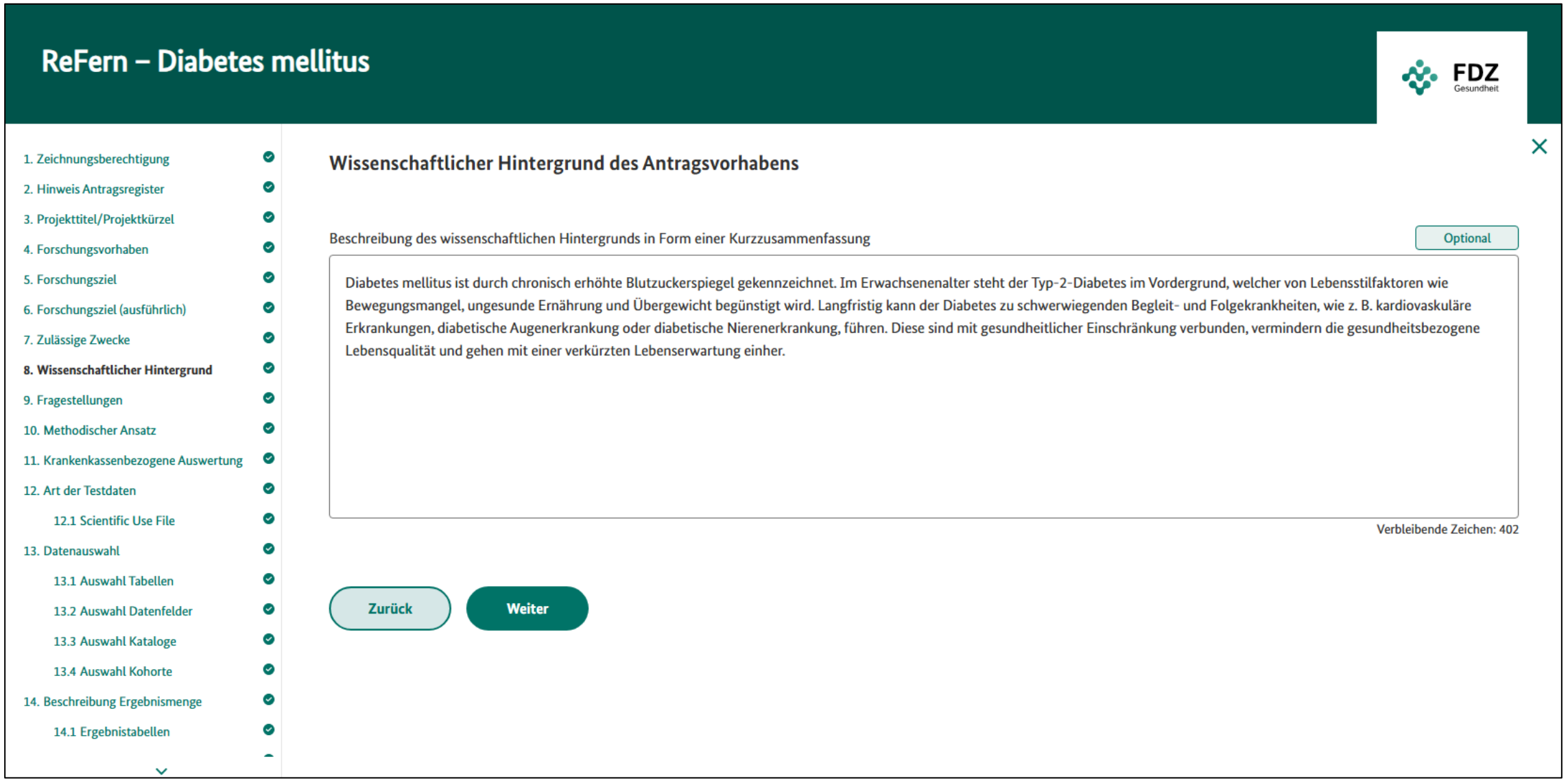
Eine Kurzbeschreibung des **Forschungsziels**, in der das Ziel der Datenverarbeitung erläutert wird.



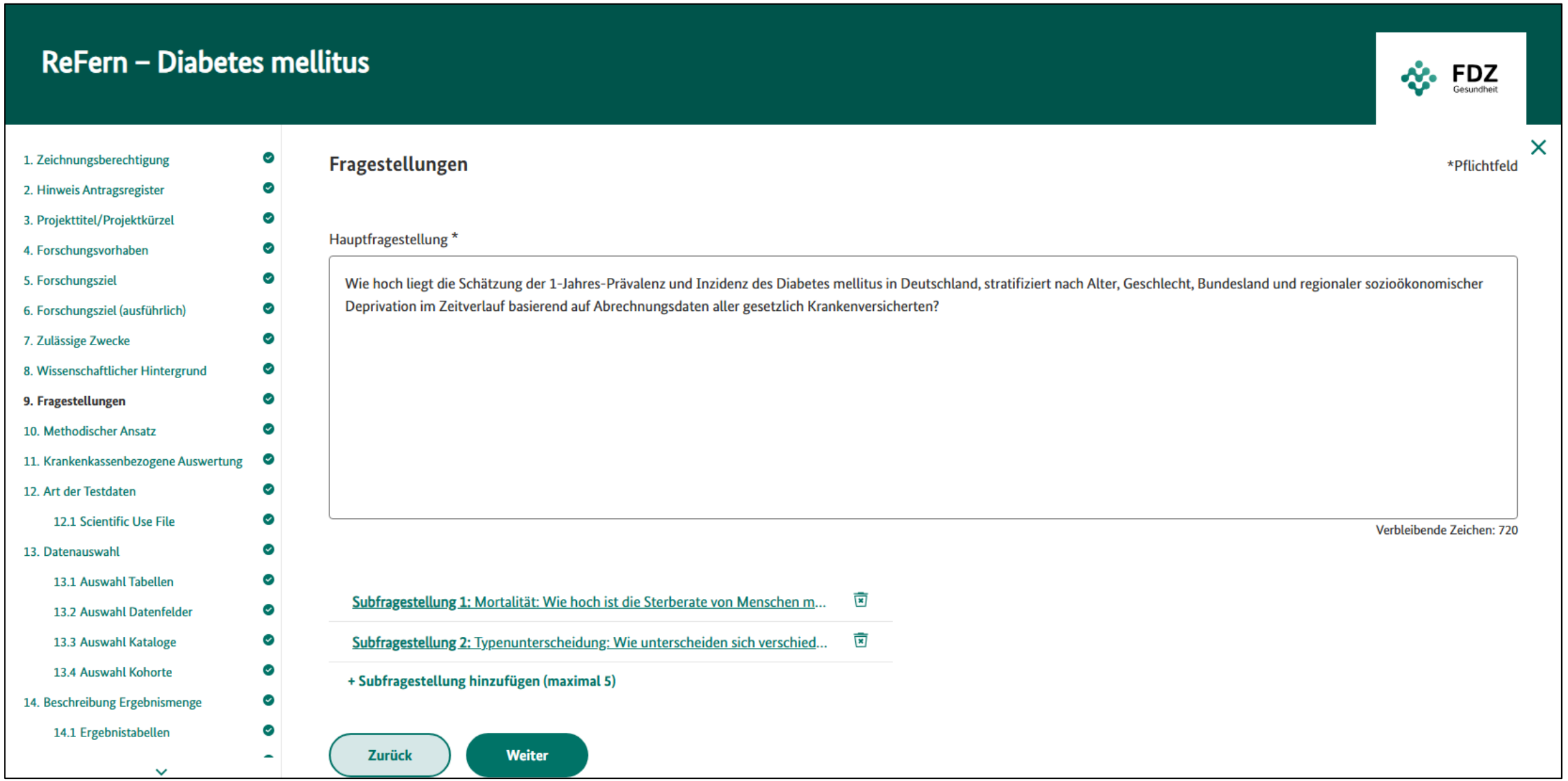
Die **ausführliche Beschreibung des Forschungsziels ist optional**, aber kann bei der Prüfung hilfreich für das Verständnis sein.



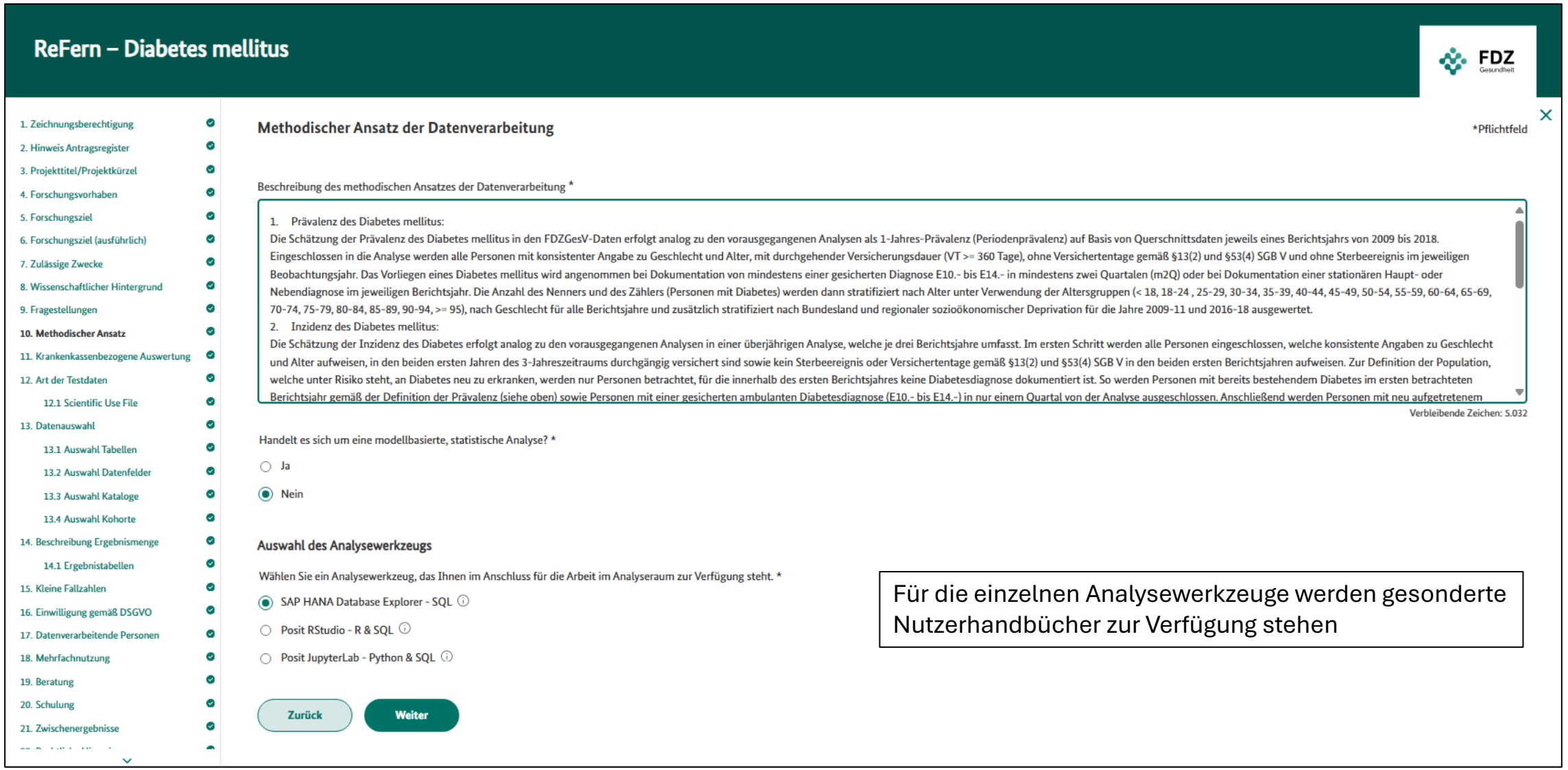
Bei den Zwecken können nur diejenigen ausgewählt werden, die auch bei der **Institutionsregistrierung** angegeben wurden.



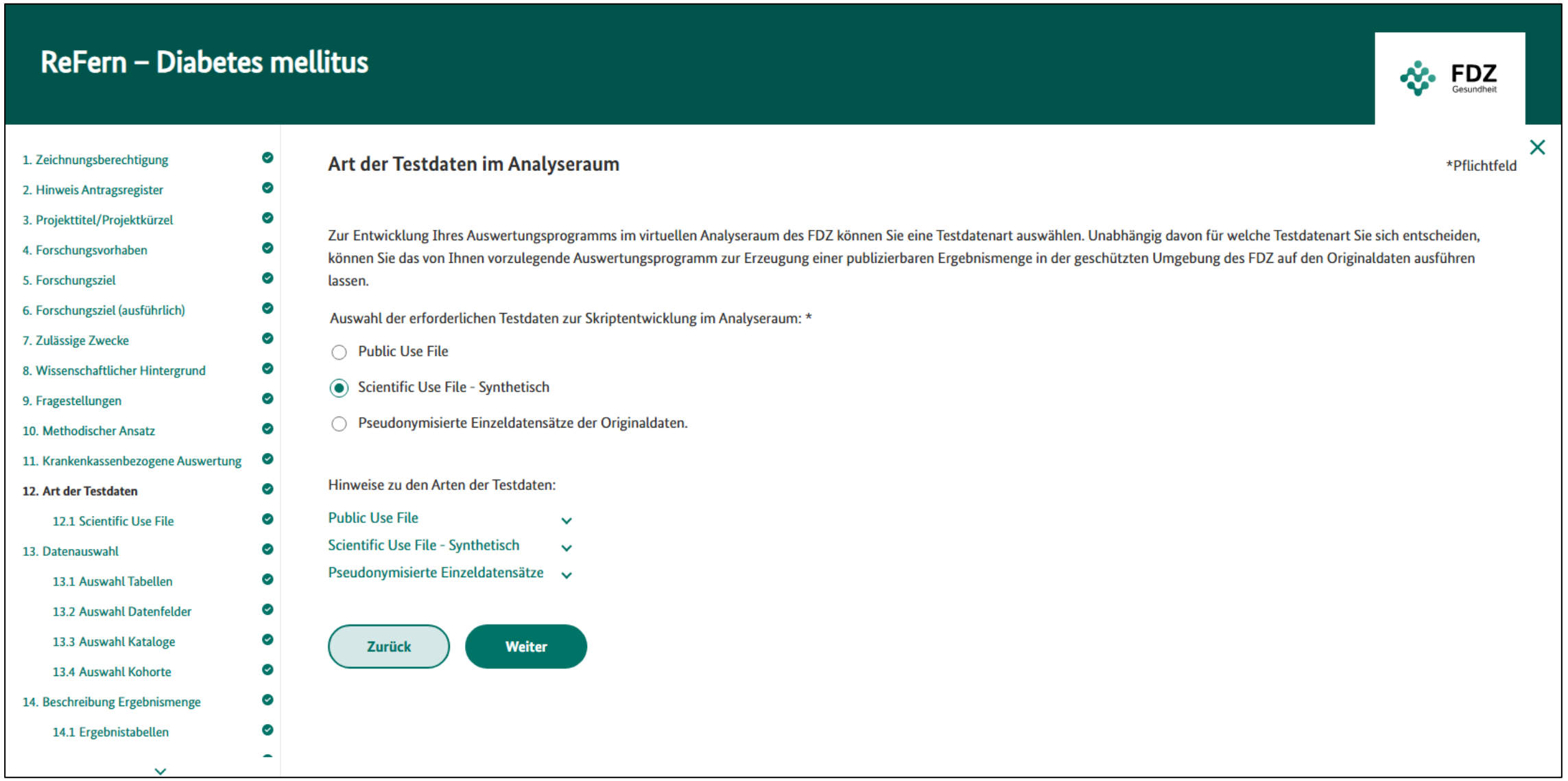
Optional kann der **wissenschaftliche Hintergrund** kurz zusammengefasst werden. Auch dies dient dem besseren Verständnis bei der Antragsprüfung.



Im Textfeld wird die **Hauptfragestellung** mit möglichen **Subfragstellungen** eingegeben. Die Fragestellung ist ein zentraler Punkt bei der Antragsprüfung und sollte daher entscheidende Informationen enthalten, ob z.B. ein bestimmter Zeitraum, Altersgruppen, Krankheiten (ICD-Codes) oder Regionen betrachtet werden sollen.



Der Methodische Ansatz ist ein weiterer zentraler Bestandteil der Prüfung, welche Daten bereitgestellt werden können. Deswegen müssen die Angaben möglichst präzise und vollständig sein. Die **Auswahl des Analysewerkzeug** entscheidet darüber, welche Programmiersprache bzw. Software **später im Analyseraum zur Verfügung steht**.



Zur Entwicklung des Auswertungsskripts im virtuellen Analyseraum stehen **mehrere Testdatentypen** zur Verfügung (Hinweise dazu auf der nächsten Seite). Unabhängig von der Auswahl wird ein Auswertungsskript zur Erzeugung einer publizierbaren Ergebnismenge in der isolierten Umgebung des FDZ Gesundheit auf den Echtdaten ausgeführt. Im allgemeinen empfiehlt das FDZ Gesundheit die Auswahl des Scientific Use File (SUF).

Hinweise zu den Arten der Testdaten:

Public Use File

Das Public Use File ist ein Beispieldatensatz zur Skriptentwicklung im Analyseraum, bei dem den Datennutzenden Daten der im Antrag ausgewählten Berichtsjahre, Tabellen und Tabellenfelder bereitgestellt werden. Es werden zufällig gezogene Daten von 3% der Versicherten (ca. 2,2 Millionen Versicherte) bereitgestellt. Univariate Verteilungen aus den Originaldaten sind abgebildet, inhaltliche Zusammenhänge zwischen Variablen, zwischen Tabellen und zwischen Berichtsjahren sind zerstört, Verknüpfungen (Joins) von Tabellen sind jedoch möglich.

Weitere Informationen

Scientific Use File - Synthetisch

Das Scientific Use File (synthetisch) ist ein Beispieldatensatz zur Skriptentwicklung im Analyseraum, bei dem den Datennutzenden Daten der im Antrag ausgewählten Berichtsjahre, Tabellen und Tabellenfelder bereitgestellt werden. Es werden Daten von maximal 1 Million zufällig gezogenen synthetischen Versicherten der im Antrag ausgewählten Kohorte bereitgestellt. Die Beziehungen zwischen den Hauptmerkmalen der realen Population sind in den synthetischen Daten reproduziert. Dabei bleiben die inhaltlichen Zusammenhänge zwischen Tabellen und zwischen Berichtsjahren erhalten.

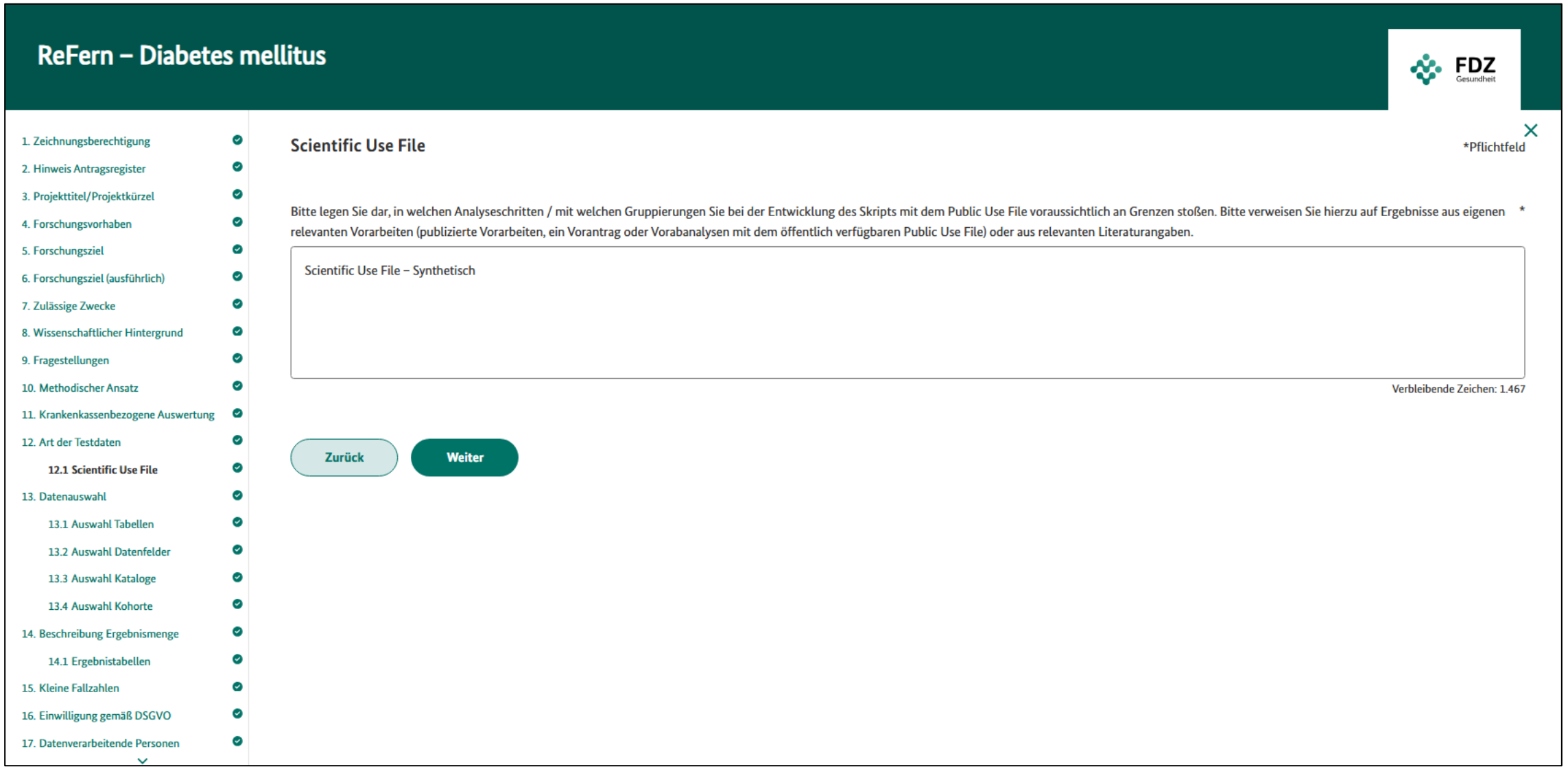
Weitere Informationen

Pseudonymisierte Einzeldatensätze

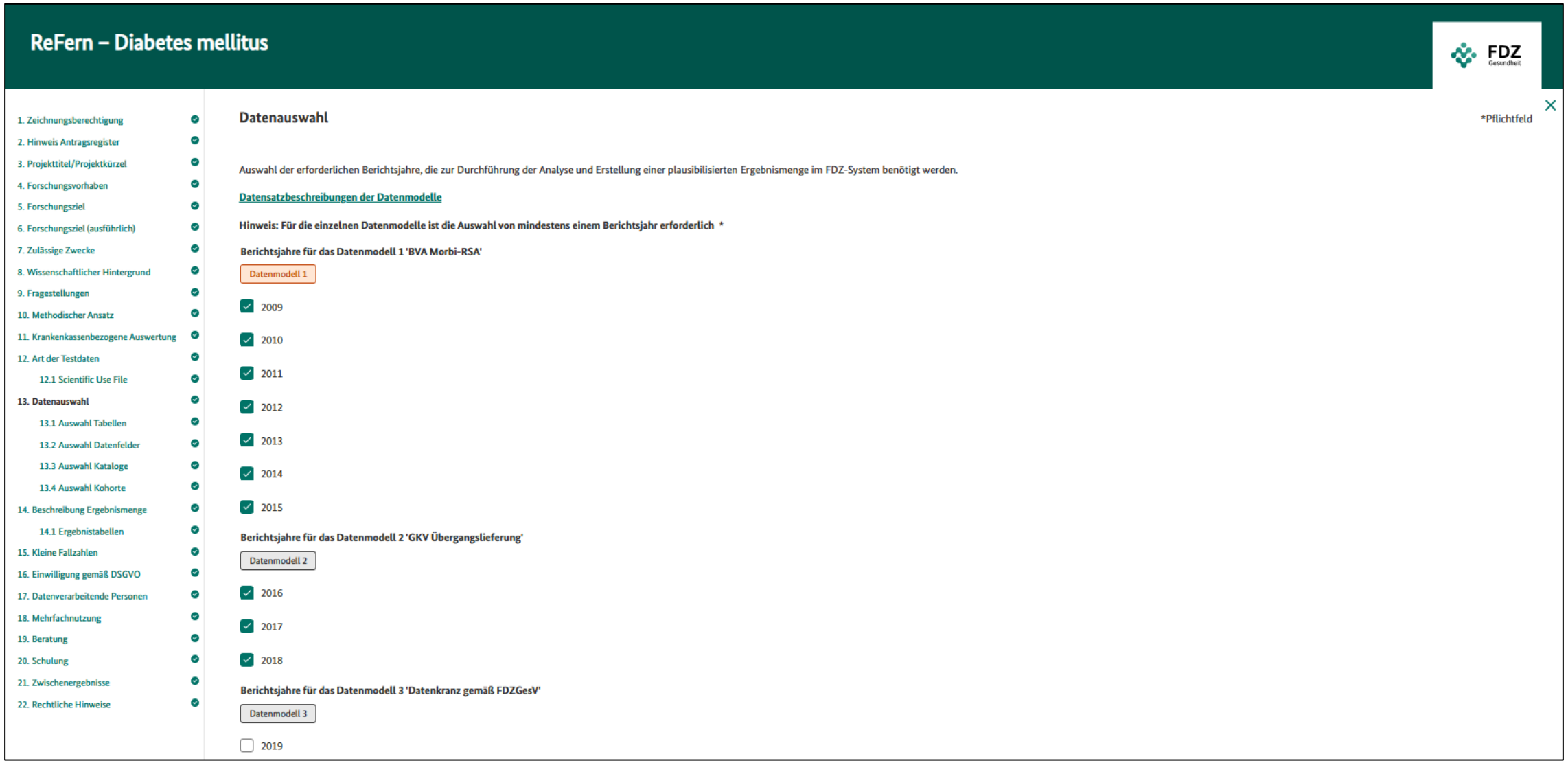
Pseudonymisierte Einzeldatensätze sind Originaldaten zur Skriptentwicklung im Analyseraum, wobei den Datennutzenden Daten der im Antrag ausgewählten Berichtsjahre, Tabellen und Tabellenfelder bereitgestellt werden. Es werden Daten von 10% (jedoch maximal 1 Million) zufällig gezogenen Versicherten der im Antrag ausgewählten Kohorte bereitgestellt. Pseudonymisierte Einzeldatensätze werden nur Personen bereitgestellt, welche einer Geheimhaltungspflicht nach § 203 StGB unterliegen. Die Nutzung von pseudonymisierten Einzeldatensätzen ist nur für Anträge möglich, die mit anderen Beispieldatensätzen nicht durchgeführt werden können, und die Erforderlichkeit muss inhaltlich begründet werden.

Link zur Public Use File:
https://zenodo.org/records/15057924

Das Public Use File ist geeignet, um schon vor der Antragstellung ein Skript zu entwickeln, damit die Arbeitszeit in den kostenpflichtigen Analyseräumen kürzer ausfällt.

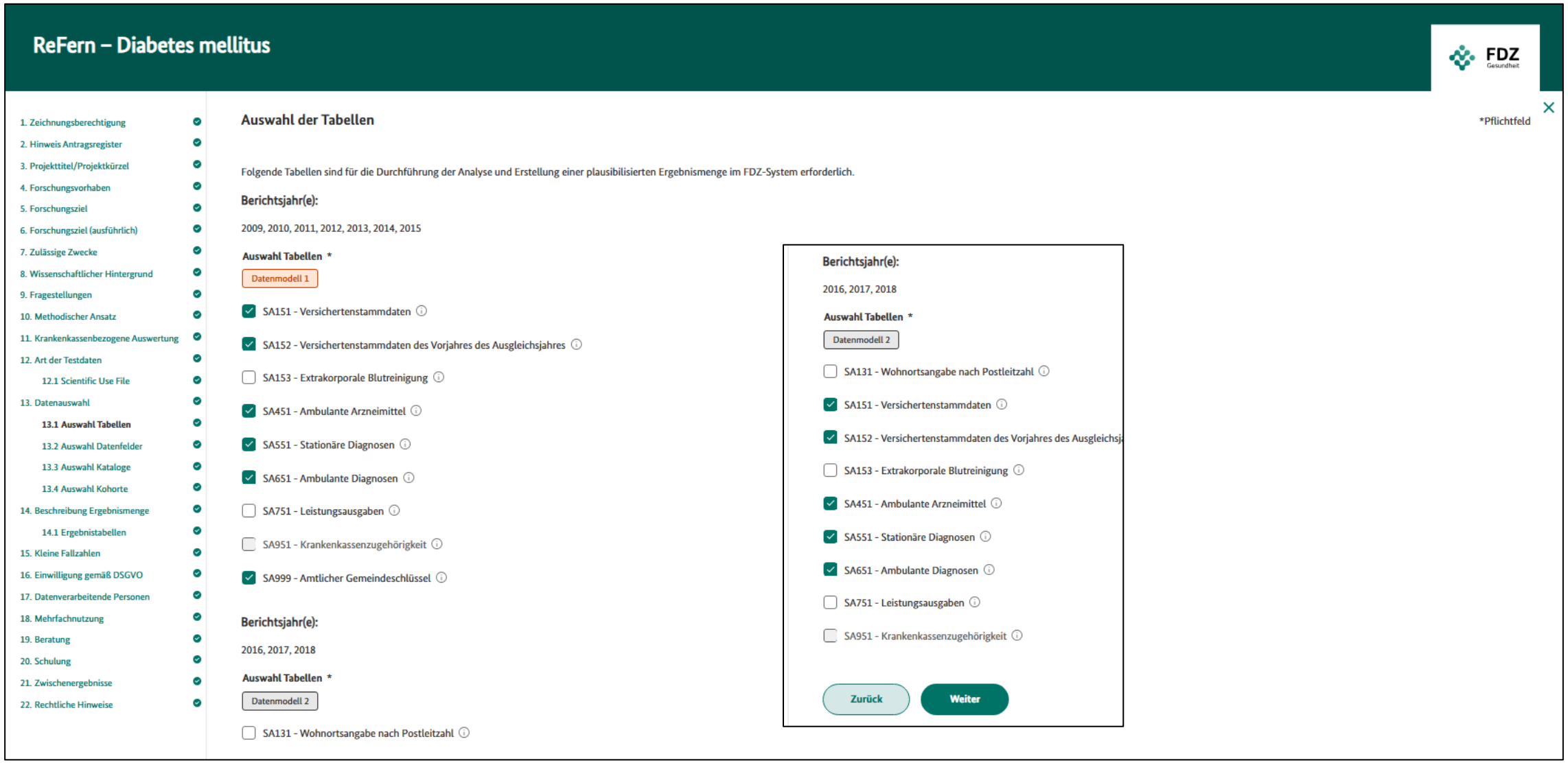


Bei der Verwendung des Public Use File (PUF) ist keine Begründung notwendig, bei der Verwendung des SUF reicht die Angabe von "Scientific Use File - Synthetisch". Bei pseudonymisierten Einzeldatensätzen (pED) ist ein **Nachweis notwendig**, warum die anderen Testdatenformen nicht ausreichen. Ein Nachweis könnte zum Beispiel ein vorheriger Antrag oder Publikation sein.

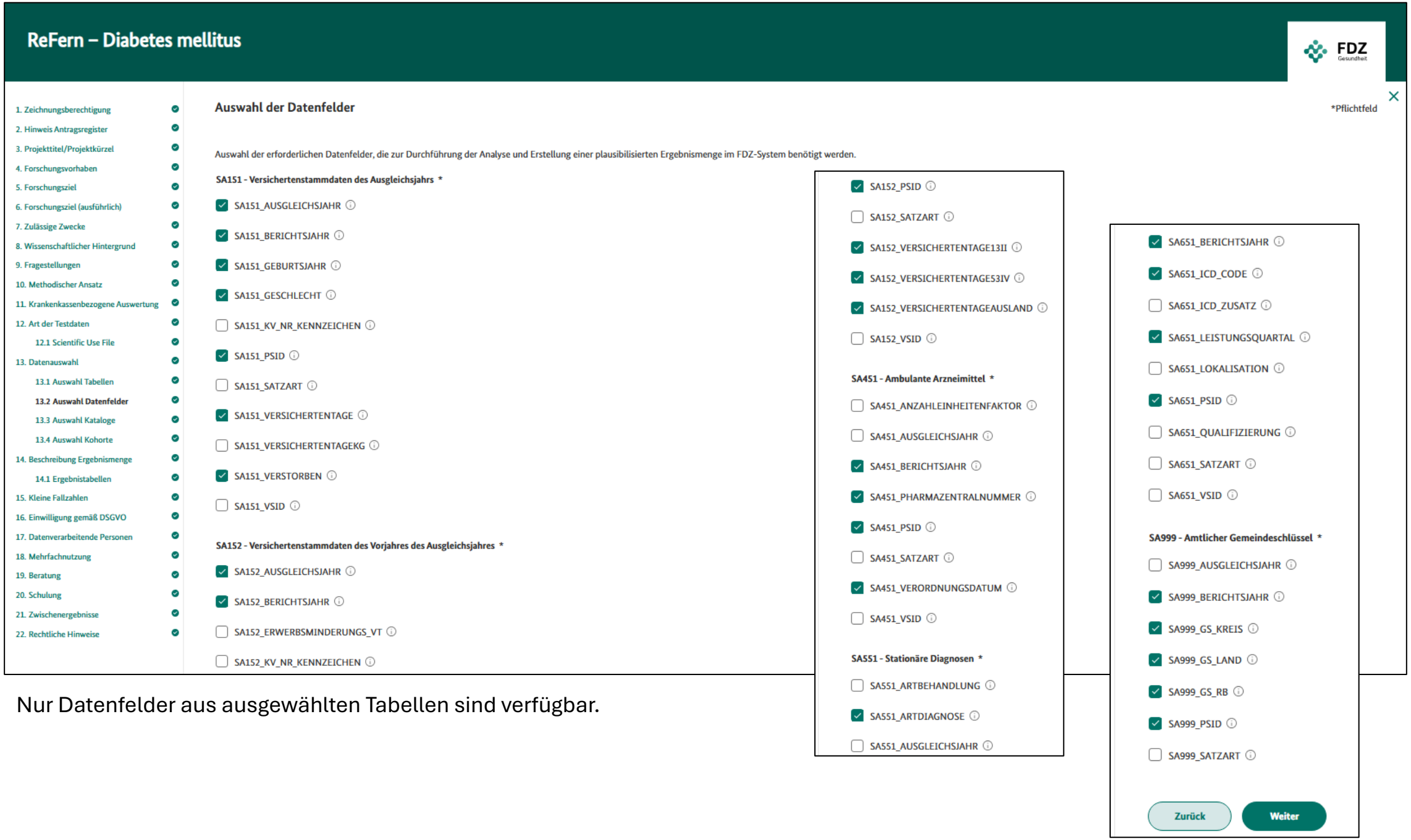


Es werden nur so viele Daten freigegeben, wie für das Projekt basierend auf der Beschreibung notwendig sind.
Datensatzbeschreibungen zu den Datenmodellen 1 (Berichtsjahre 2009-2015), 2 (2016-2018) und 3 (ab 2019) sind hier zu finden:

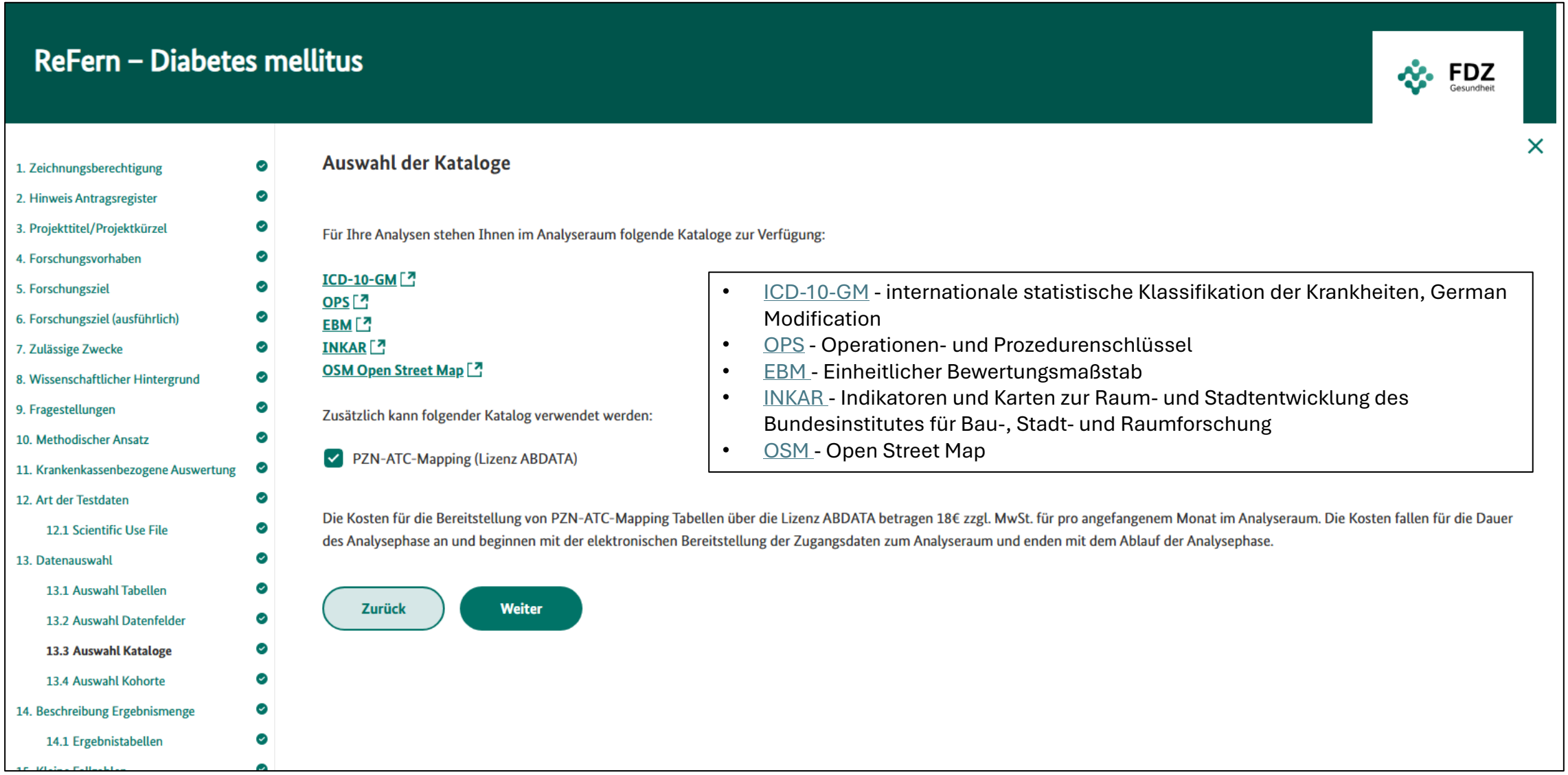
- https://fdz-gesundheit.github.io/datensatzbeschreibung_fdz_gesundheit/
- https://github.com/FDZ-Gesundheit/datensatzbeschreibung_fdz_gesundheit/tree/main



Die verfügbaren Tabellen hängen davon ab, welche **Berichtsjahren** mit dem jeweiligen Datenmodell ausgewählt wurden.

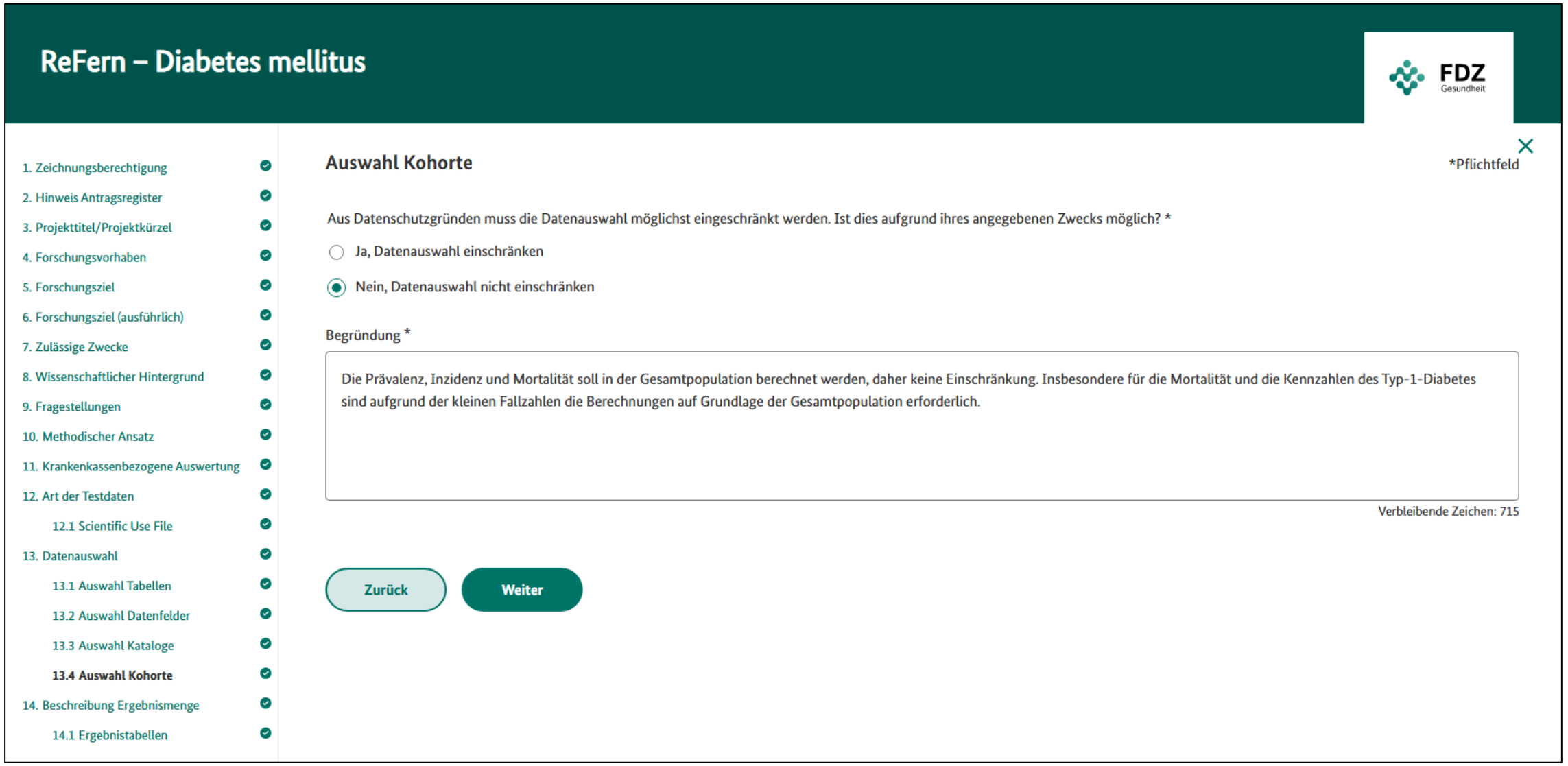


Nur Datenfelder aus ausgewählten Tabellen sind verfügbar.

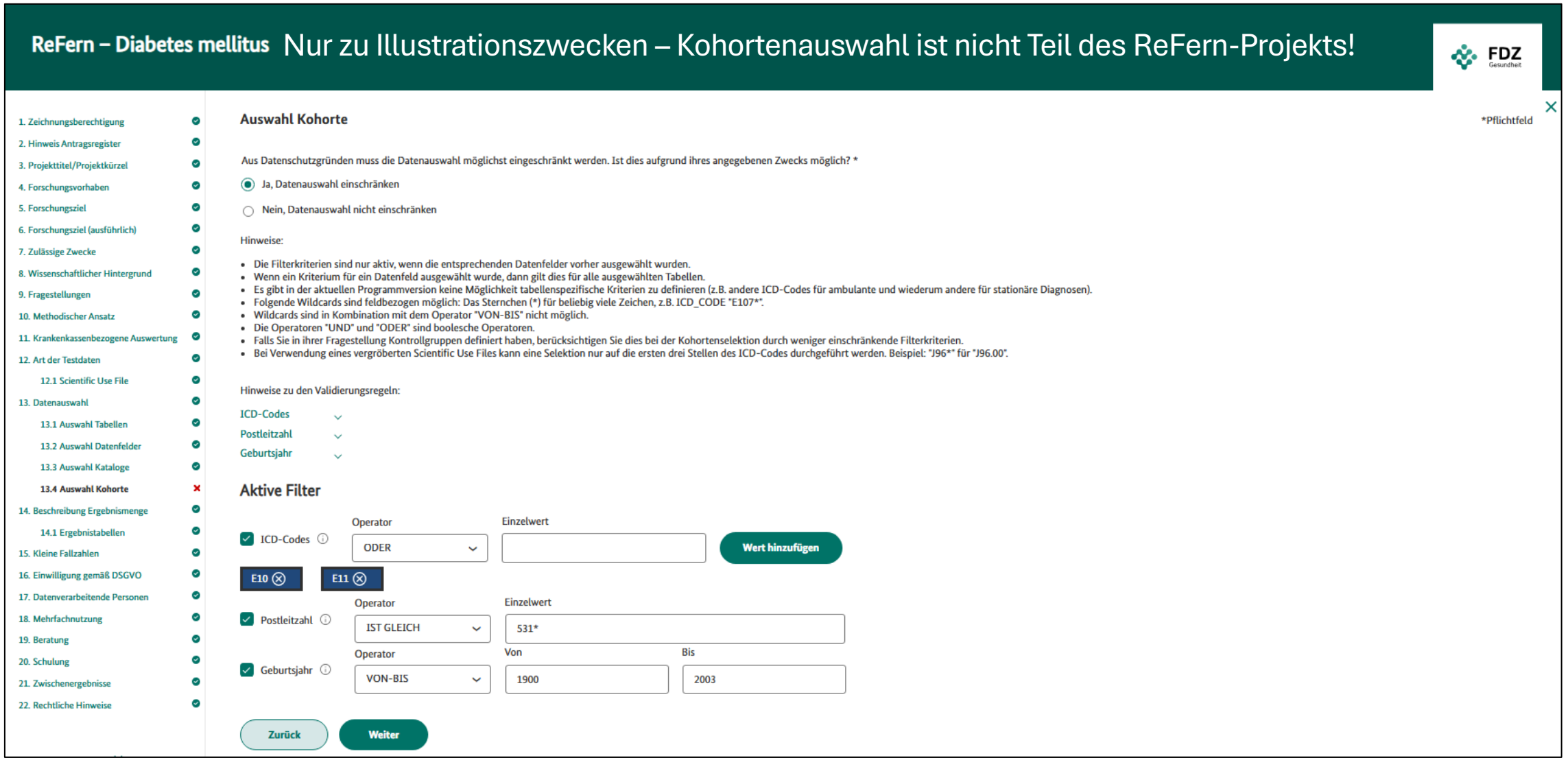


- ICD-10-GM - internationale statistische Klassifikation der Krankheiten, German Modification
- OPS - Operationen- und Prozedurenschlüssel
- EBM - Einheitlicher Bewertungsmaßstab
- INKAR - Indikatoren und Karten zur Raum- und Stadtentwicklung des Bundesinstitutes für Bau-, Stadt- und Raumforschung
- OSM - Open Street Map

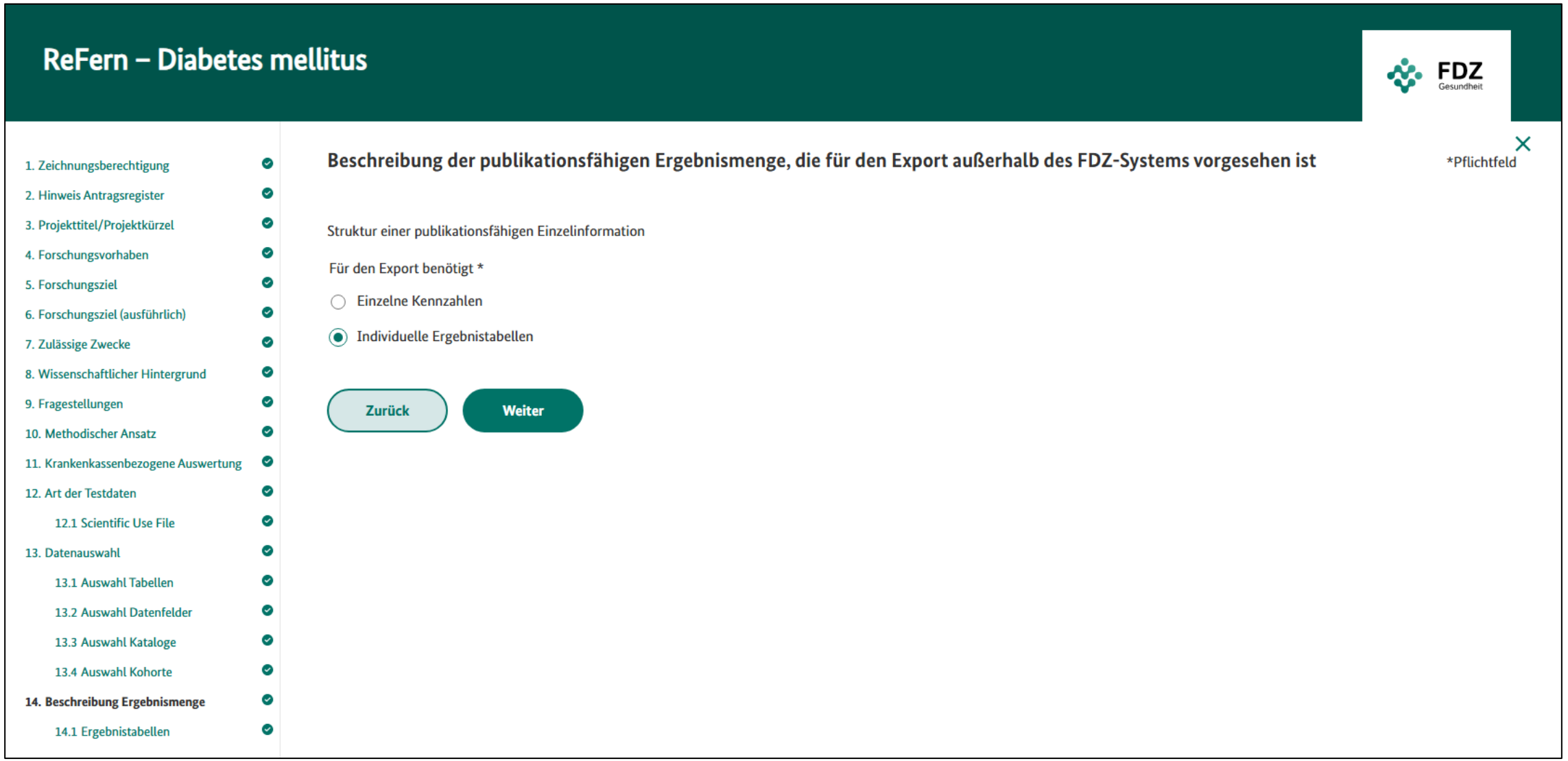
Für einen Datenlinkage stehen mehrere Kataloge kostenlos zur Verfügung (Stand 2025). Ein Zugriff auf den ABDATA-Katalog für ein PZN-ATC-Mapping kann ohne gesonderte Lizenz bereitgestellt werden kann, er ist aber kostenpflichtig.



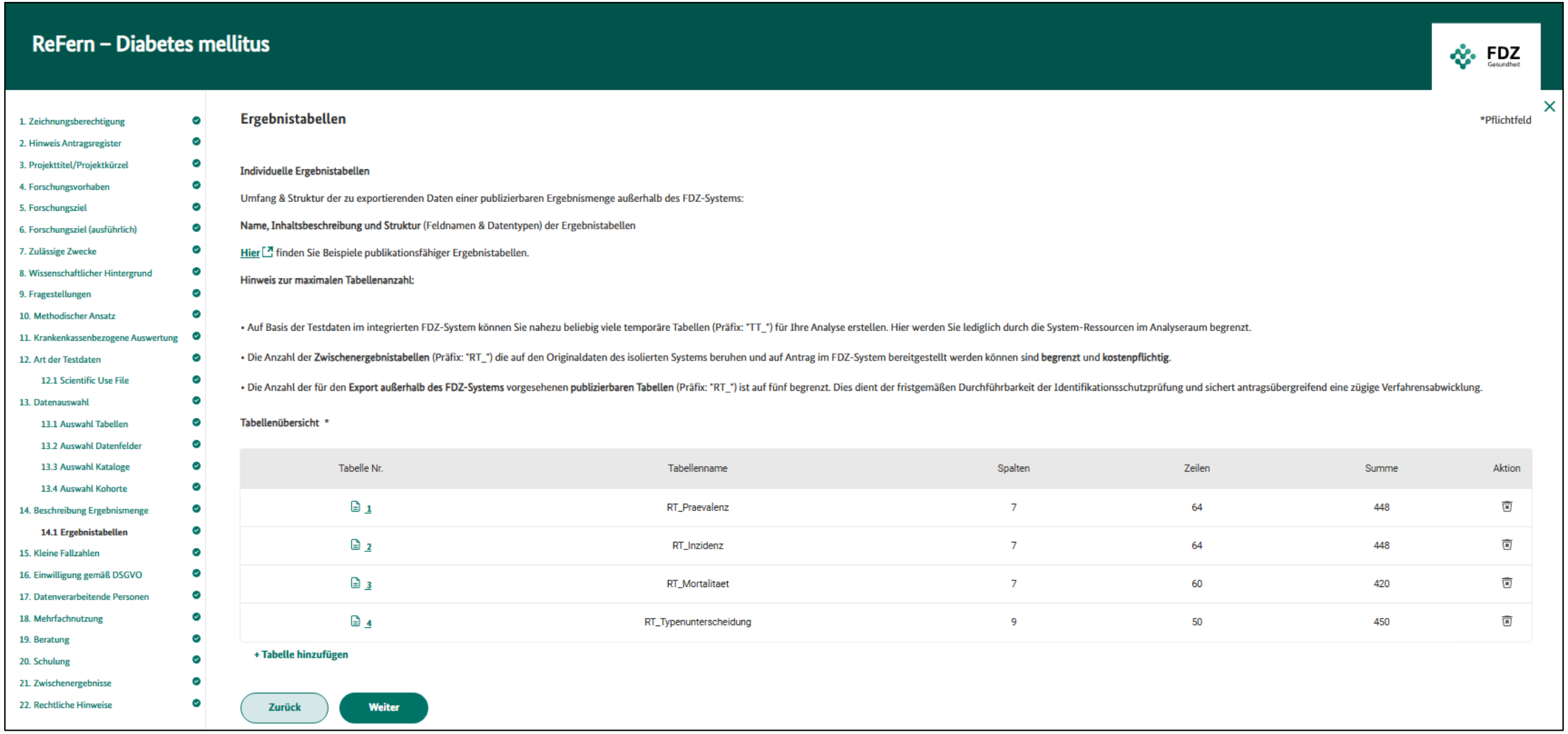
Die Datenauswahl muss auf **eine Kohorte** beschränkt werden, wenn als Testdatentyp SUF oder pED ausgewählt wurde und für das Projekt nur bestimmte Krankheiten (ICD-Codes), Regionen (Postleitzahl) oder Geburtsjahrgänge untersucht werden. **Falls eine Beschränkung nicht möglich ist, muss das begründet werden.**



Beispiel für Kohortenauswahl. Durch die ausgewählten Filter enthält die Kohorte nur Versicherte mit Diabetes mellitus Typ 1 (E10) oder Typ 2 (E11), die im Bonner Raum wohnen (Postleitzahl beginnt mit 531) und im Berichtsjahr 2021 älter als 18 Jahre alt waren.



Am Ende der Analysephase läuft das Skript auf den Echtdaten und erzeugt eine Ergebnismenge. Bevor diese Ergebnismenge für den Download und damit dem Export aus dem FDZ Antragsportal freigegebenen wird, prüfen Mitarbeitende des FDZ Gesundheit, ob die Ergebnismenge dem Identifikationsschutz gerecht wird. Bei dieser Prüfung wird unter anderem untersucht, ob die Ergebnismenge mit dem übereinstimmt, was beantragt wurde.



Der Inhalt von jeder Tabelle und deren Spalten muss beschrieben werden. Es können bis zu **5 Ergebnistabellen mit maximal 10 Spalten** beantragt werden. Eine Tabelle darf maximal 2400 Zellen enthalten, die eine Anzahl an Versicherten oder darauf basierende Berechnungen angeben. Beschreibende Spalten, die die Versicherten charakterisieren, werden bei der Berechnung der Zellzahl nicht berücksichtigt. So würde zum Beispiel bei einer Tabelle mit den Spalten "Geschlecht", "Alter", "ICD-Code" und "CNT_D_PSID" (=Anzahl Versicherte) nur die Spalte "CNT_D_PSID" zu der Berechnung Maximalanzahl an erlaubten Zellen beitragen.

Ausschnitt einer Tabellenbeschreibung

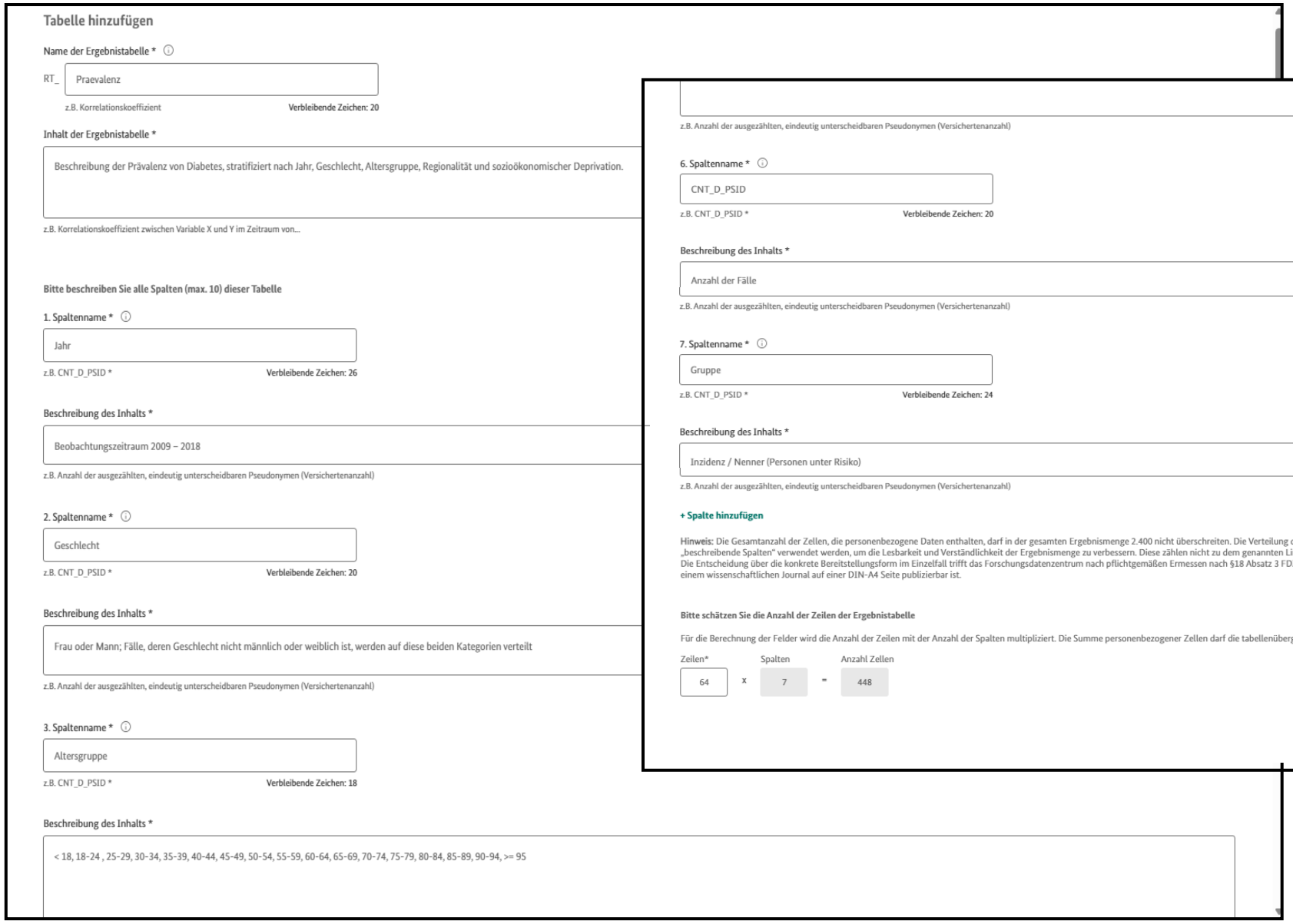


Tabelle hinzufügen

Name der Ergebnistabelle *

RT_ Praevalenz

z.B. Korrelationskoeffizient — Verbleibende Zeichen: 20

Inhalt der Ergebnistabelle *

Beschreibung der Prävalenz von Diabetes, stratifiziert nach Jahr, Geschlecht, Altersgruppe, Regionalität und sozioökonomischer Deprivation.

z.B. Korrelationskoeffizient zwischen Variable X und Y im Zeitraum von...

Bitte beschreiben Sie alle Spalten (max. 10) dieser Tabelle

1. Spaltenname *

Jahr

z.B. CNT_D_PSID * — Verbleibende Zeichen: 26

Beschreibung des Inhalts *

Beobachtungszeitraum 2009 – 2018

z.B. Anzahl der ausgezählten, eindeutig unterscheidbaren Pseudonymen (Versichertenanzahl)

2. Spaltenname *

Geschlecht

z.B. CNT_D_PSID * — Verbleibende Zeichen: 20

Beschreibung des Inhalts *

Frau oder Mann; Fälle, deren Geschlecht nicht männlich oder weiblich ist, werden auf diese beiden Kategorien verteilt

z.B. Anzahl der ausgezählten, eindeutig unterscheidbaren Pseudonymen (Versichertenanzahl)

3. Spaltenname *

Altersgruppe

z.B. CNT_D_PSID * — Verbleibende Zeichen: 18

Beschreibung des Inhalts *

< 18, 18-24 , 25-29, 30-34, 35-39, 40-44, 45-49, 50-54, 55-59, 60-64, 65-69, 70-74, 75-79, 80-84, 85-89, 90-94, >= 95

z.B. Anzahl der ausgezählten, eindeutig unterscheidbaren Pseudonymen (Versichertenanzahl) — Verbleibende Zeichen: 1.958

Spalte löschen

6. Spaltenname *

CNT_D_PSID

z.B. CNT_D_PSID * — Verbleibende Zeichen: 20

Beschreibung des Inhalts *

Anzahl der Fälle

z.B. Anzahl der ausgezählten, eindeutig unterscheidbaren Pseudonymen (Versichertenanzahl) — Verbleibende Zeichen: 1.984

Spalte löschen

7. Spaltenname *

Gruppe

z.B. CNT_D_PSID * — Verbleibende Zeichen: 24

Beschreibung des Inhalts *

Inzidenz / Nenner (Personen unter Risiko)

z.B. Anzahl der ausgezählten, eindeutig unterscheidbaren Pseudonymen (Versichertenanzahl) — Verbleibende Zeichen: 1.959

Spalte löschen

+ Spalte hinzufügen

Hinweis: Die Gesamtanzahl der Zellen, die personenbezogene Daten enthalten, darf in der gesamten Ergebnismenge 2.400 nicht überschreiten. Die Verteilung dieses Limits auf die einzelnen Tabellen (maximal 5 Tabellen) kann beliebig erfolgen. Darüber hinaus können sogenannte „beschreibende Spalten" verwendet werden, um die Lesbarkeit und Verständlichkeit der Ergebnismenge zu verbessern. Diese zählen nicht zu dem genannten Limit.
Die Entscheidung über die konkrete Bereitstellungsform im Einzelfall trifft das Forschungsdatenzentrum nach pflichtgemäßen Ermessen nach §18 Absatz 3 FDZGesV. Gemäß der Nebenbestimmungen ist eine publizierbare Ergebnistabelle eine Tabelle, die in ihrem Umfang und ihrer Struktur in einem wissenschaftlichen Journal auf einer DIN-A4 Seite publizierbar ist.

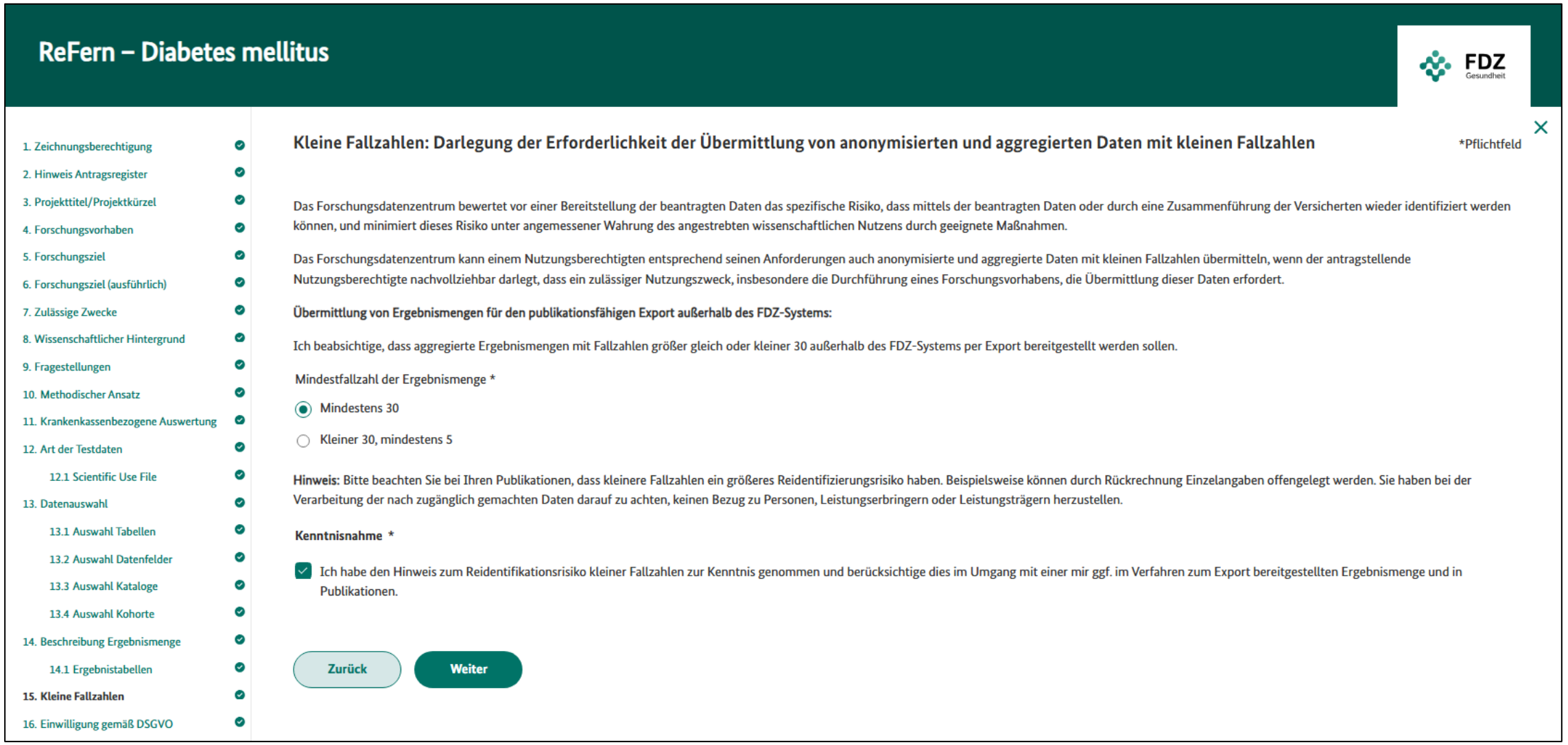
Bitte schätzen Sie die Anzahl der Zeilen der Ergebnistabelle

Für die Berechnung der Felder wird die Anzahl der Zeilen mit der Anzahl der Spalten multipliziert. Die Summe personenbezogener Zellen darf die tabellenübergreifende Gesamtzahl an Zellen nicht überschreiten.

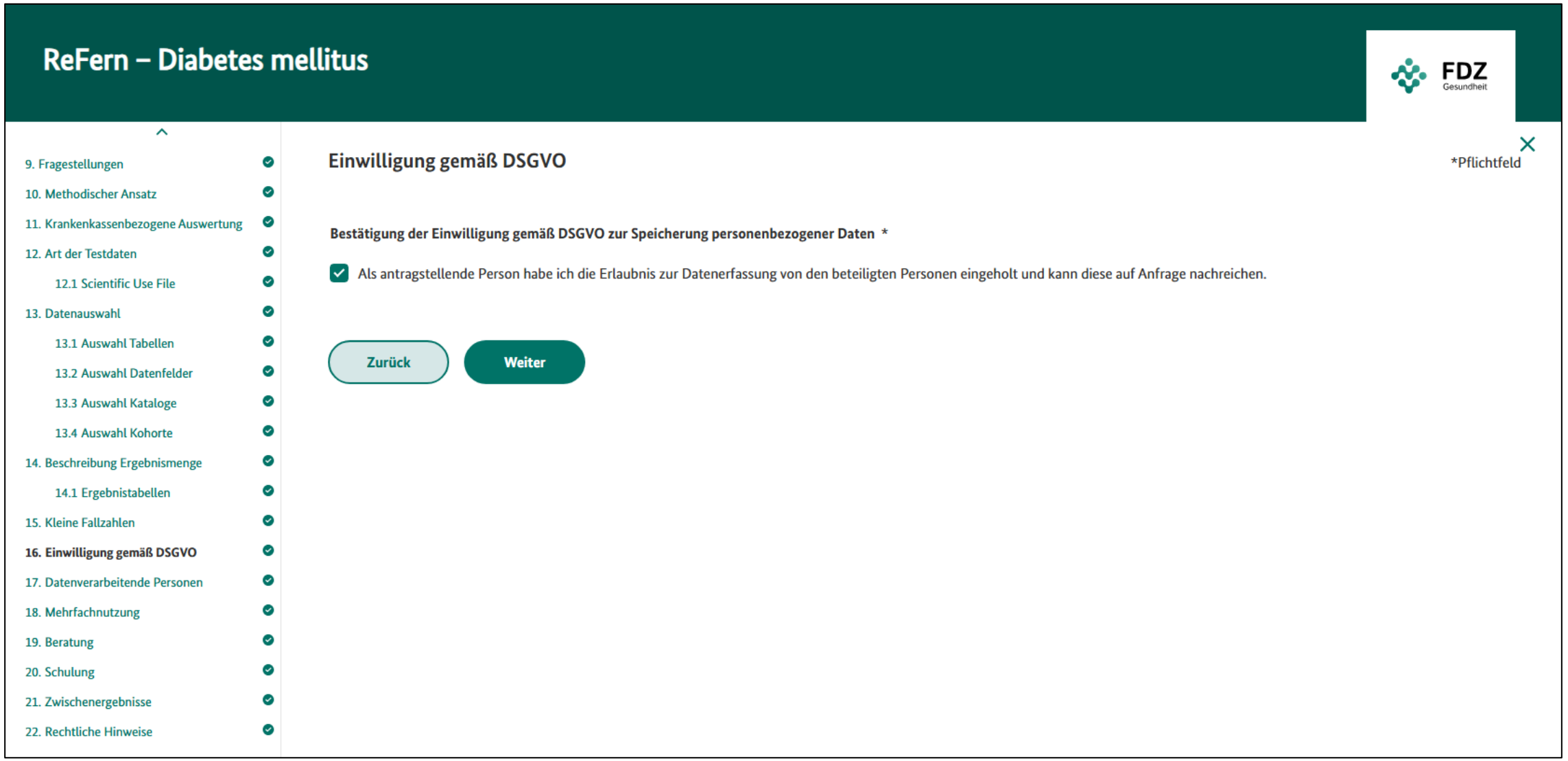
Zeilen*		Spalten		Anzahl Zellen
64	x	7	=	448

Übernehmen | Abbrechen

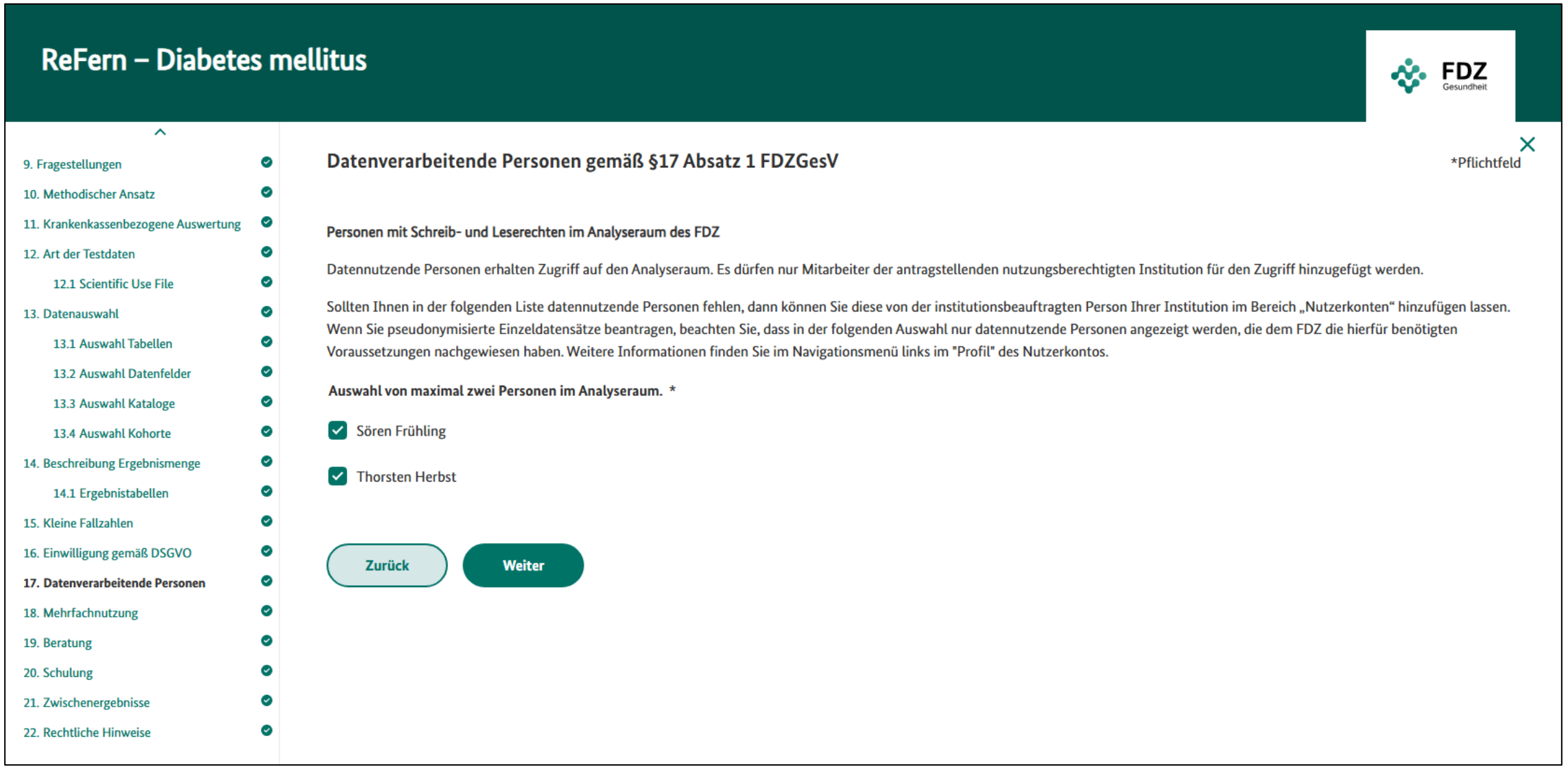
Der FDZ-Datensatz enthält neben „weiblich" und „männlich" auch „divers" (Geschlecht 3) und „keine Angabe" (Geschlecht 4). Aufgrund geringer Fallzahlen können diese Kategorien die Mindestfallzahl unterschreiten. Forschende entscheiden eigenständig über den Umgang damit, etwa durch zufällige Zuordnung zu binären Geschlechtern oder Ausschluss – jeweils aus rein statistischen Gründen.



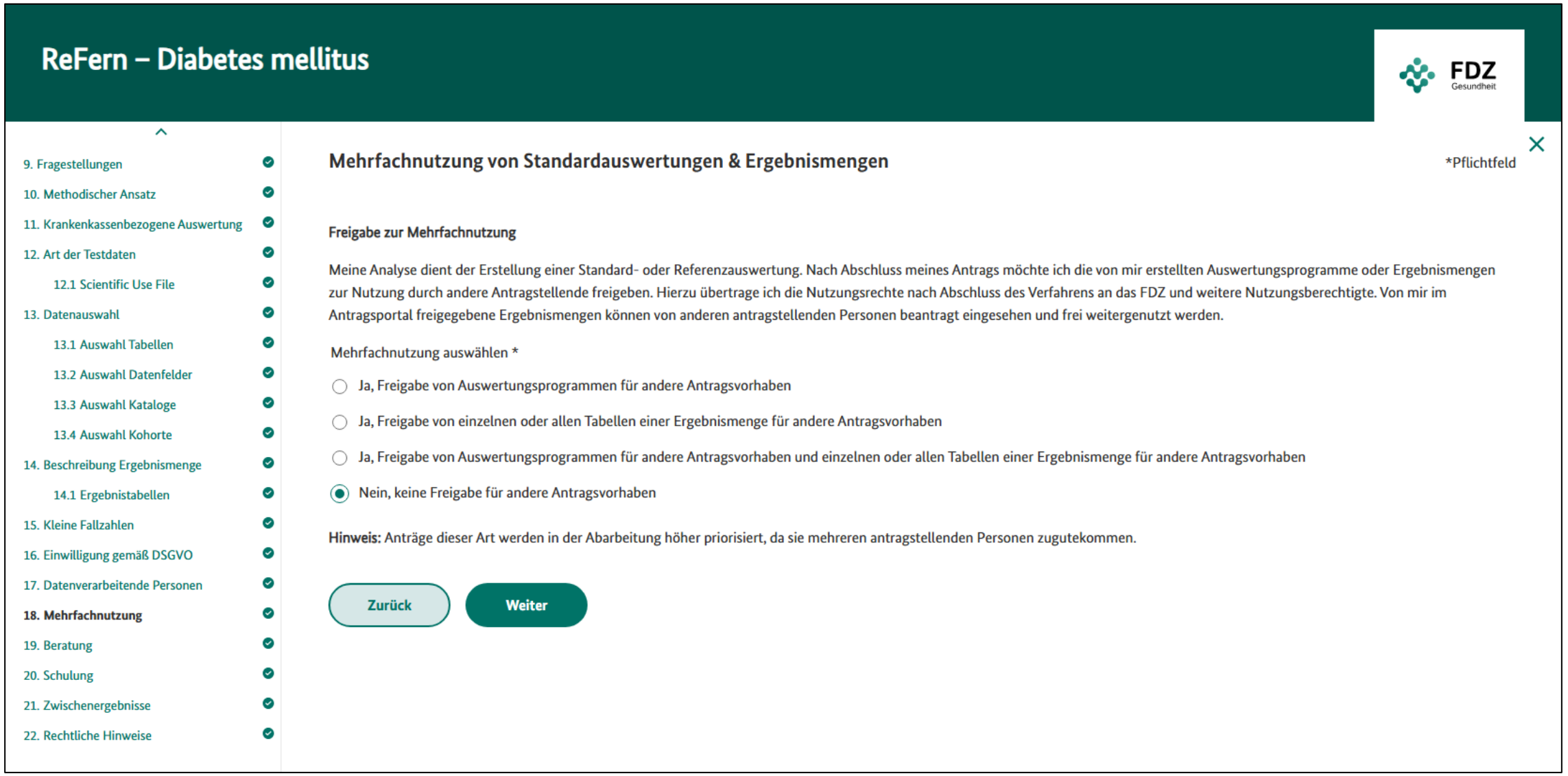
Die **Mindestfallzahl liegt bei 30.** Das heißt, jeder personenbezogene Wert bezieht sich auf mindestens 30 Versicherte. Eine Absenkung ist auf Beantragung möglich, doch lehnt das FDZ Absenkungen in der Regel ab, wenn die geplante Auswertung zwei oder mehr dieser Kriterien erfüllt: Feine regionale Granularität, wenn Geschlecht 3 und 4 nicht ausgeschlossen werden, feine Granularität des Alters, Krankenkassen-spezifische Auswertung oder feine Granularität von Erkrankungen (feiner als ICD-3-Steller).



Bei der Antragerstellung und Datenverarbeitung werden von allen Beteiligten personenbezogene Daten erfasst und gespeichert.



Nur datenverarbeitende Personen erhalten Zugang zu den Analyseräume und entwickeln das Analyseskript. Die antragstellende Person kann gleichzeitig auch datennutzende Person sein, muss es aber nicht.



Standardauswertungen sind derzeit technisch noch nicht möglich.

ReFern – Diabetes mellitus

FDZ Gesundheit

9. Fragestellungen
10. Methodischer Ansatz
11. Krankenkassenbezogene Auswertung
12. Art der Testdaten
    12.1 Scientific Use File
13. Datenauswahl
    13.1 Auswahl Tabellen
    13.2 Auswahl Datenfelder
    13.3 Auswahl Kataloge
    13.4 Auswahl Kohorte
14. Beschreibung Ergebnismenge
    14.1 Ergebnistabellen
15. Kleine Fallzahlen
16. Einwilligung gemäß DSGVO
17. Datenverarbeitende Personen
18. Mehrfachnutzung
19. Beratung
20. Schulung
21. Zwischenergebnisse
22. Rechtliche Hinweise

Beratung

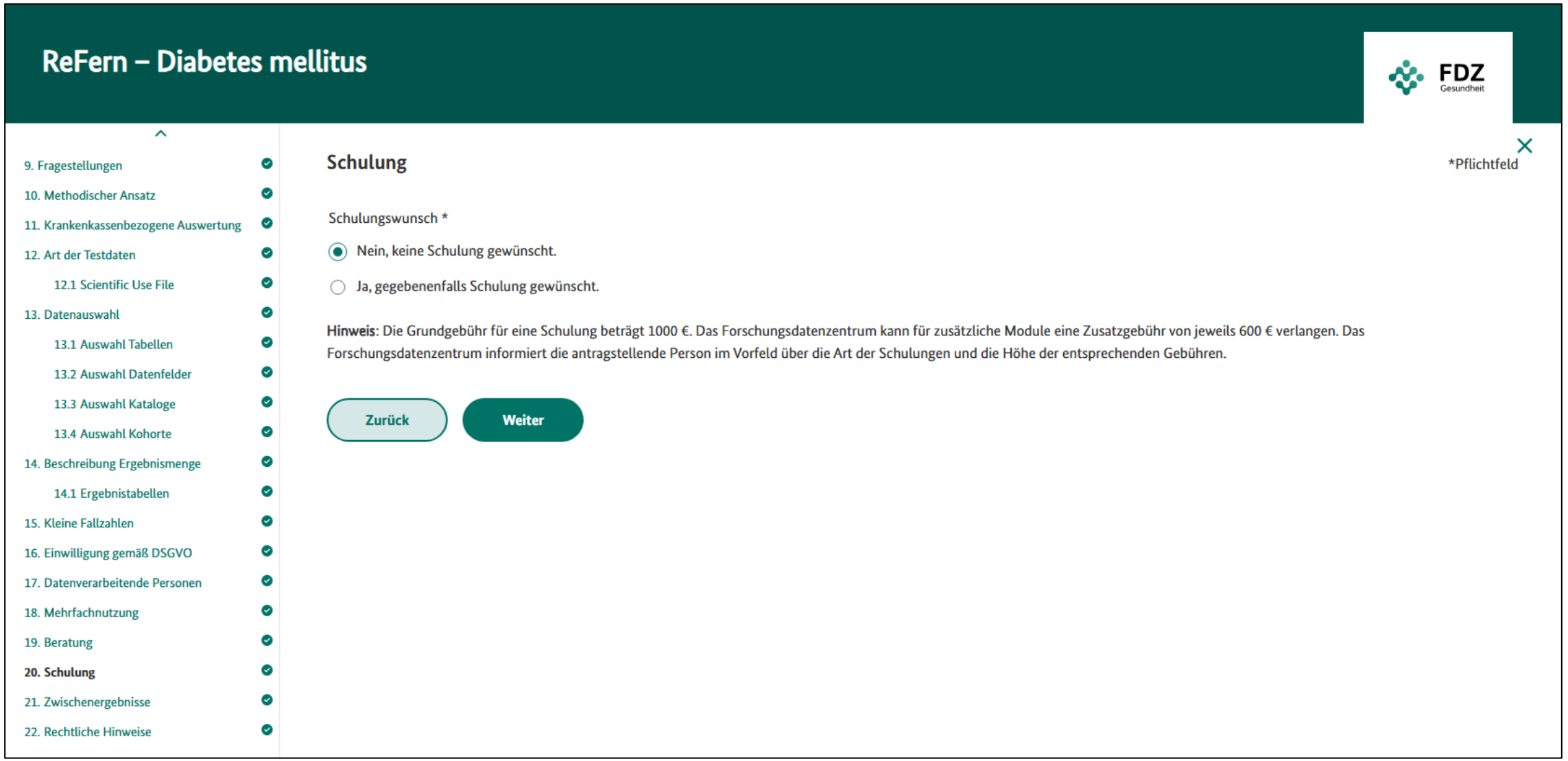
*Pflichtfeld

Beratungswunsch *

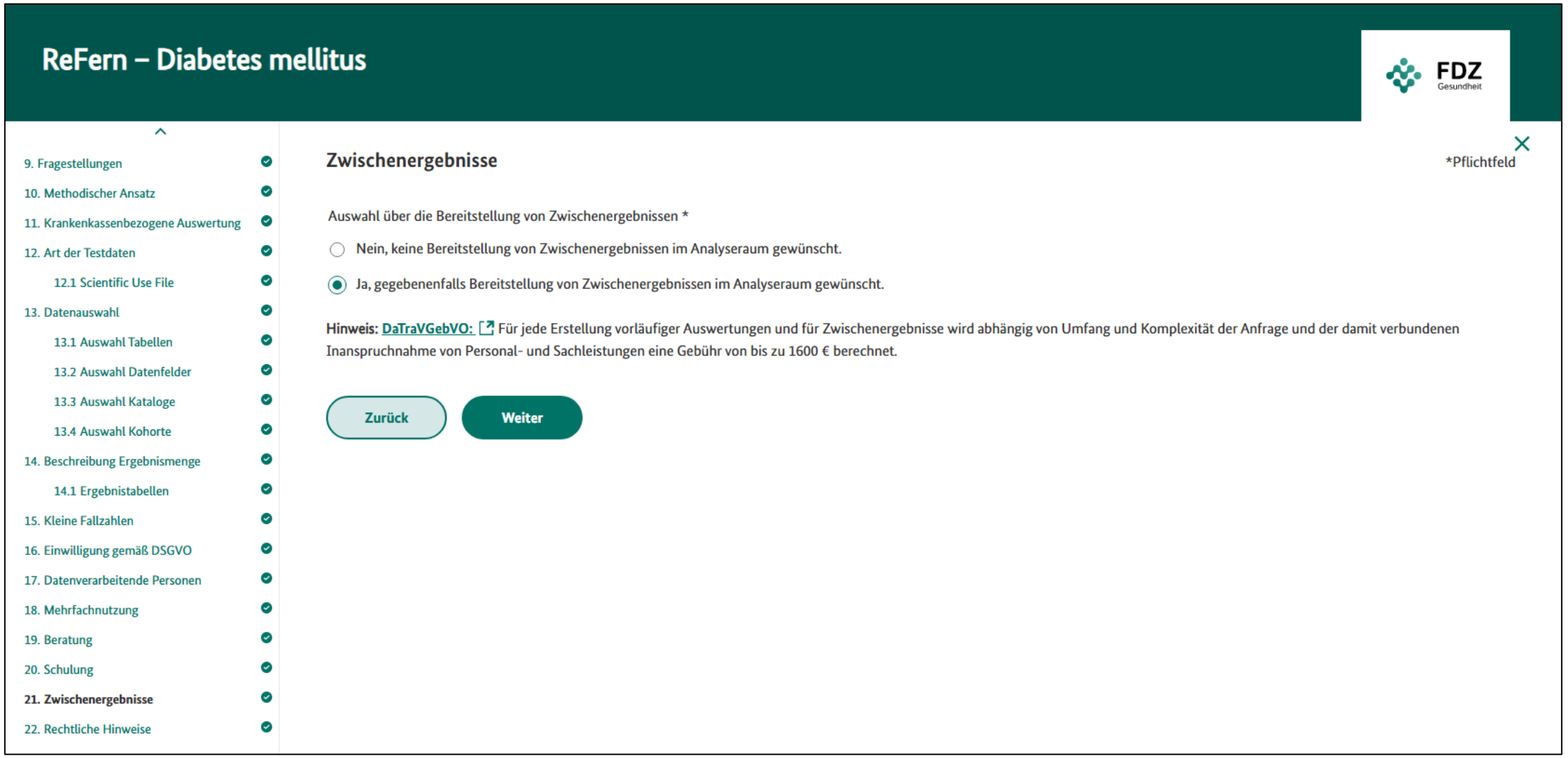
- (●) Nein, keine Beratung gewünscht.
- ( ) Ja, gegebenenfalls Beratung gewünscht.

Hinweis: §18 Absatz 2 FDZGesV, §7 DaTraGebV Die Beratung soll auf acht Stunden pro Antrag begrenzt werden. In begründeten Ausnahmefällen kann die Beratungszeit auf bis zu 16 Stunden pro Antrag erweitert werden. Für jede Beratung wird abhängig von Umfang und Komplexität der Anfrage und der damit verbundenen Inanspruchnahme von Personal- und Sachleistungen eine Gebühr von 50 € bis **maximal** 1600 € berechnet. **Die Kosten für Beratungsleistungen rechnen wir halbstündlich ab, somit entstehen für Sie Kosten in Höhe von 50 € pro angefangene halbe Stunde.** Bitte beachten Sie, dass Beratungsleistungen von mehr als 16 Stunden nicht zulässig sind.

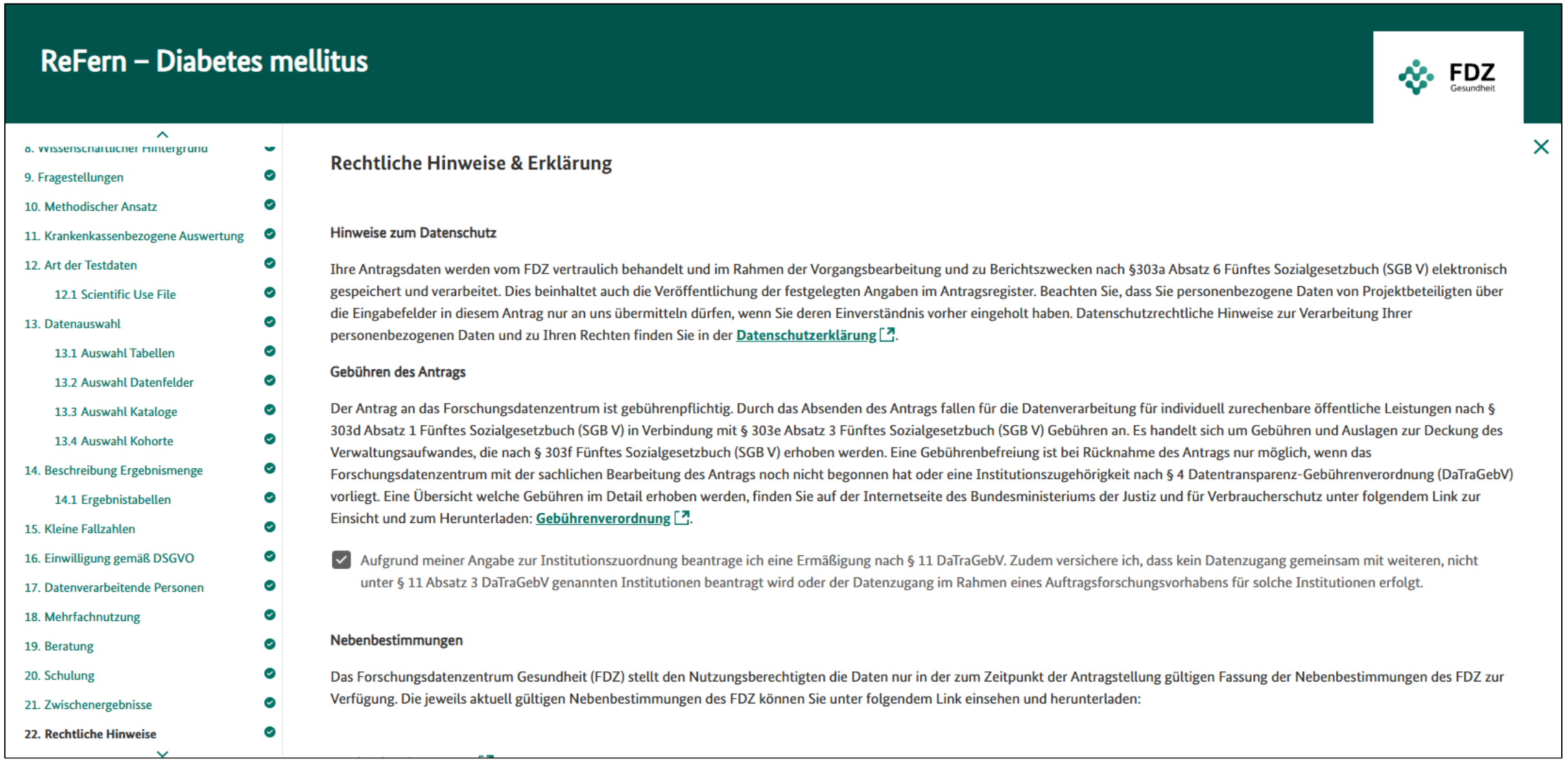
Zurück | Weiter



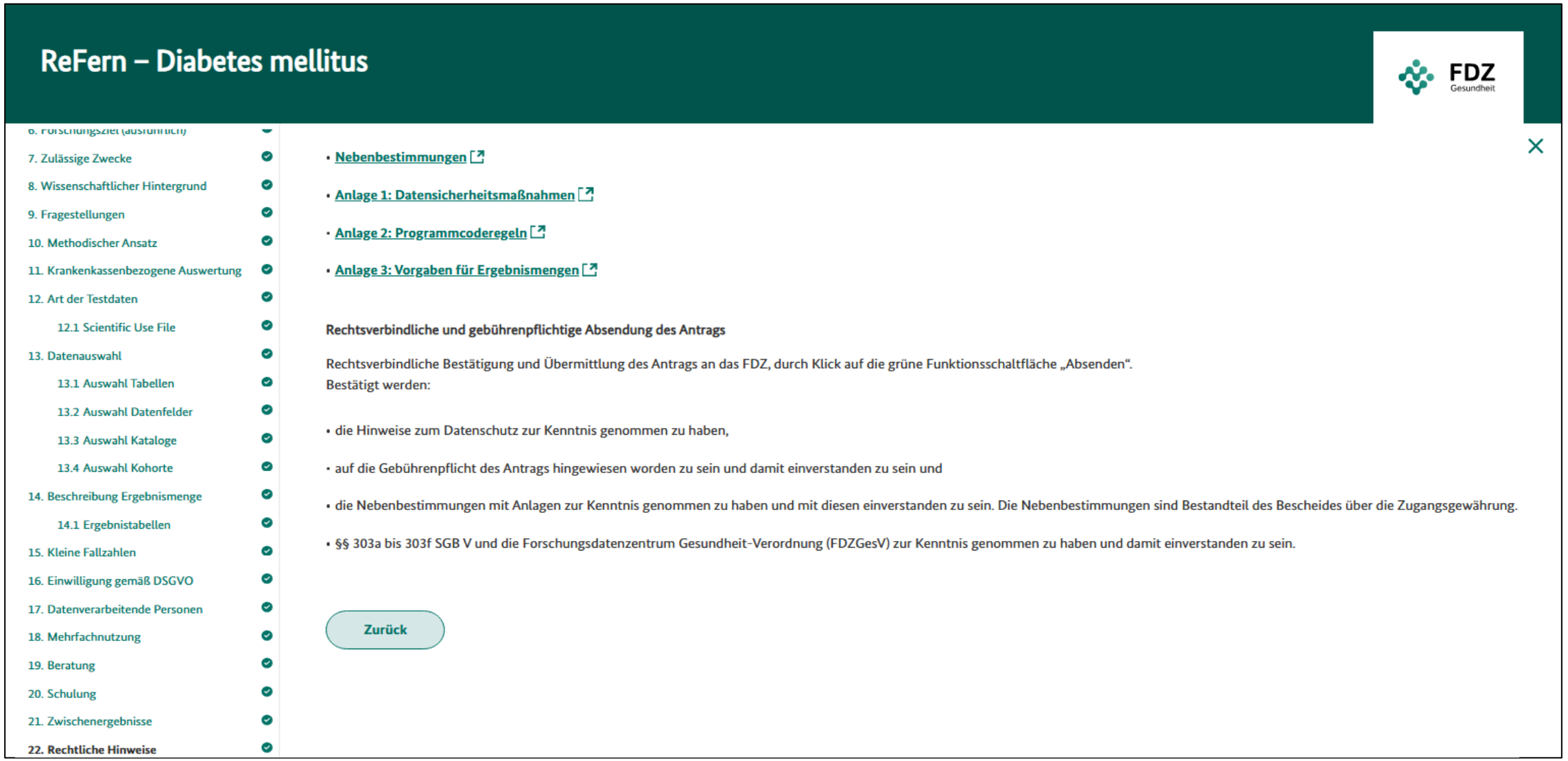
Für aufwändigere Anfragen zu verschiedenen Themen kann eine kostenpflichtige Schulung gebucht werden.



Bis zu fünfmal pro Antrag können kostenpflichtige Zwischenergebnisse im Analyseraum bereitgestellt werden. Es ist aber nicht verpflichtend, Zwischenergebnisse anzufordern, wenn hier "Ja" angegeben wird.



Teil 1 der rechtlichen Hinweise.



Teil 2 der rechtlichen Hinweise. Zu den Nebenbestimmungen gehören drei Anlagen.
Nach Betätigung des „Absenden“-Buttons gelangt der Antragsentwurf zur zeichnungsberechtigten Person, die den Antrag zur Prüfung freigibt oder zur Korrektur an antragstellende Person zurückschickt.

Gebührenverordnung

Die Datentransparenz-Gebührenverordnung (DaTraGebV) legt fest, welche Gebühren das FDZ Gesundheit erheben muss. Beispielhaft werden sie hier für den Demo-Antrag berechnet:

Position	DaTraGebV	Einzelposition	Anzahl	Gebühr	Erläuterung
Grundgebühr	§ 5	4.000,00€	1	4.000,00€	
Zusatzgebühr für Jahrgänge	§ 6 Abs. 2	1.000,00€	10	10.000,00€	Zehn Berichtsjahre (2009-2018)
Bereitstellung Daten in gesicherter virtueller Umgebung	§ 6 Abs. 3 1000 € pro Tag	1.000,00€	8	8.000,00€	Allgemein gibt es kein Limit bei der Anzahl der Tage im Analyseraum (außer bei pEDs). Für das Beispiel hier wurde eine Fertigstellung des Skripts nach 8 Tagen angenommen.
Erstellung von Zwischenergebnissen	§ 7	109,00€	1	109,00€	Einmal wurden Zwischenergebnisse angefordert, um die Sinnhaftigkeit der Ergebnisse zu prüfen.
Zwischensumme				**22.109,00€**	
Rechnungsbetrag nach Gebührenermäßigung auf ein Zehntel	§11 Abs. 3 u. 4	Ermäßigung wurde beantragt und bewilligt		**2.210,90€**	Das RKI ist eine Institution der Gesundheitsberichterstattung des Bundes und ist für eine Gebührenermäßigung berechtigt. Dies gilt u.a. auch für Hochschulen und Hochschulkliniken. Liste der berechtigten Institutionen unter §11 DaTraGebV Abs. 3 (Link)