

CCP: Datenschutznachtrag zur föderierten Analyse

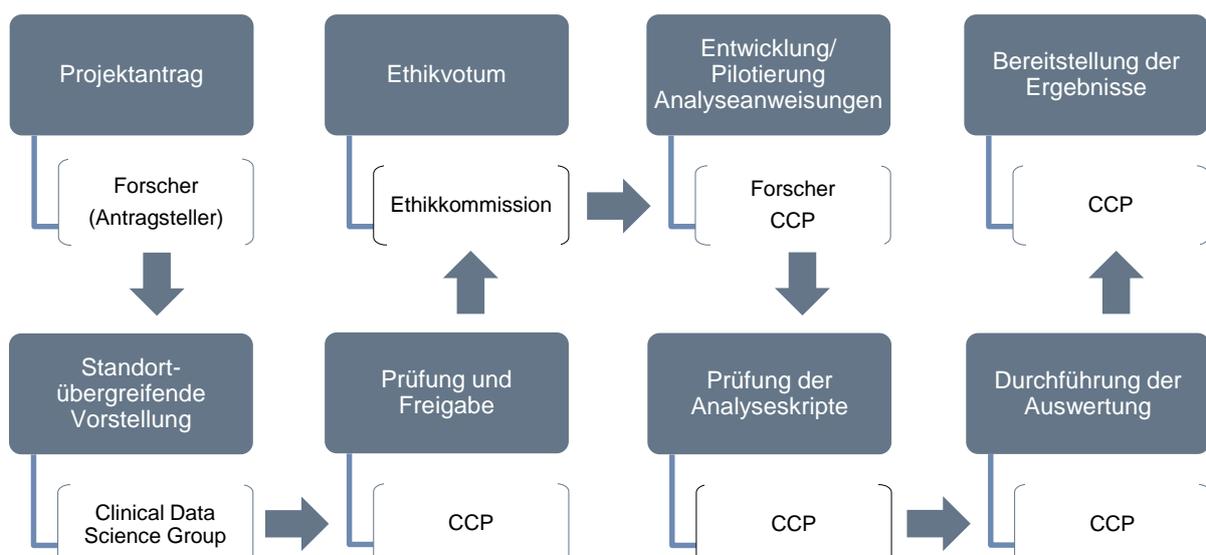
Stand: 20.06.2023

Version	Datum	Beschreibung	Autoren
1.0	20.06.2023	Erste Fassung	M. Lambarki, T. Kussel, D. Maier, B. Uhl, S. Meyer, D. Juárez, M. Lablans

Dieses Dokument stellt einen Datenschutznachtrag zum aktuellen Datenschutzkonzept der CCP-IT vom 23.11.2020 dar. Diese Erweiterung beschreibt Erweiterungen zur Erfüllung der Nutzeranforderungen¹ zur Datenauswertung und wird nach Evaluation der Verfahren in der Praxis in das Hauptdokument des Datenschutzkonzepts integriert.

Föderierte Analyse

Mit der Föderierten Analyse können Forschende des DKTK und dessen Kooperationspartner von den Standorten genehmigte Projekte den Datenbestand der Brückenköpfe für projektspezifische Fragestellungen über die CCP auswerten. Im Unterschied zu klassischen Datenexporten, die an einer einzelnen Stelle zusammengeführt und dort ausgewertet werden, werden bei föderierten Analysen die Analyseanweisungen in die jeweiligen Brückenköpfe übermittelt, dort lokal ausgeführt und nur Ausführungsergebnisse verlassen den Standort (Grundsatz: „Nicht die Daten gehen zur Analyse, sondern die Analyse geht zu den Daten“). Die Nutzung föderierter Analysetechniken erfolgt in der CCP über ein abgestimmtes, mehrstufiges Verfahren mit einem gemeinsamen, über mehrere Standorte verteilten Kontrollmechanismus. Dies wird im Folgenden beschrieben.



¹ Zur besseren Lesbarkeit wird an manchen Stellen die gleichzeitige Verwendung männlicher und weiblicher Sprachformen verzichtet. Es wird das generische Maskulinum verwendet, wobei beide Geschlechter gleichermaßen gemeint sind.

Schritt 1: Beantragung und Prüfung des Projekts

Zunächst stellt der Forscher Informationen über das beabsichtigte Projekt zur Verfügung. Diese umfassen u.a. eine Beschreibung der beantragten Kohorte, der erforderlichen Datenpunkte, Angaben zum wissenschaftlichen Hintergrund, zur Hypothese/Zielsetzung, zu den Methoden, zu den Projektpartnern und zur Laufzeit.

Das Projekt wird in der DKTK Clinical Data Science Group² mit den Vertretern anderer potentiell teilnehmender Standorte fachlich abgestimmt. Gemäß der lokal geltenden Nutzungsordnungen der zu beteiligenden Standorte, muss die föderierte Analyse ausgewählter Daten an den Standorten beantragt werden. Die zentralen Einheiten der CCP unterstützen bei all diesen Prozessen, etwa durch Beratung und Prüfung der Projektangaben auf Vollständigkeit, Widersprüchlichkeit und Machbarkeit.

Schritt 2: Vorbereitung der Auswertung (interaktiv, an einem Standort)

Sodann erhält der Forscher in Zusammenarbeit mit der CCP die Möglichkeit, seine verteilte Analyse an einem ersten Standort (meist dem eigenen) vorzubereiten. Dieser Schritt soll sicherstellen, dass die Methodik und die Algorithmen geeignet sind, um das beantragte Projekt durchzuführen und die Analysekripte fehlerfrei und datenschutzrechtlich unbedenklich sind, bevor sie an die weiteren beteiligten Standorte verteilt werden.

Dieser Entwicklungs- und Erprobungsschritt erfolgt stets interaktiv in der Brückenkopf-Umgebung. Zur Anwendung kommen entweder synthetische Testdaten oder – falls bereits die lokal geltenden Nutzungsordnungen erfüllt sind – auch Echtdateien des jeweiligen Standorts. Auch bei diesem lokalen Zugriff wird, soweit es die Fragestellung zulässt, der Daten- und Funktionsumfang kontrolliert und alle Zugriffe protokolliert.

Ergebnis dieses Schrittes sind technische Analysenanweisungen inklusive der erforderlichen Auswertungsalgorithmen und -methoden. Diese kann als ein Software-Container (z.B. Docker-Container), als Skript (z.B. in R) oder in Form von Anweisungen für eine bestimmte Plattform für föderierte Analysen (wie z.B. DataSHIELD) sein. Diese Verfahren werden vom Brückenkopf an jedem Standort auf einheitliche Weise unterstützt.

Schritt 3: Prüfung an anderen Pilotstandorten (nicht interaktiv, an mehreren Standorten)

Die so erstellte Analyseanweisung wird an mind. zwei weitere Standorte verteilt und dort probenhalber ausgeführt. So kann die Korrektheit der Auswerteskripte von zwei weiteren Personen überprüft werden (Sechs-Augen-Prinzip), und auch mögliche Probleme bedingt durch die Heterogenität der Daten können auf diese Weise adressiert werden.

Schritt 4: Durchführung der Auswertungen

Die nunmehr überprüften Analyseanweisungen werden durch die CCP zur Ausführung an alle teilnehmenden Standorte versandt und dort ausgeführt. Diese Auswertung erfolgt nicht-interaktiv und ihre Ausgabe beinhaltet lediglich die zuvor gemeinsam festgelegten Analysen der verfügbaren Daten von Tumorpatienten, ohne dass Dritte die Daten selbst einsehen oder weitere Zugriffe erhalten mussten.

Die CCP überprüft die technisch aggregierten/anonymisierten Ergebnisse und übermittelt sie an den Antragsteller.

² Siehe <https://dktk.dkfz.de/klinische-plattformen/dkdk-clinical-data-science-group>