

Hinweise zum Antragsformular zur Bereitstellung pseudonymisierter Forschungsdaten

Mit diesen Hinweisen möchten wir Sie beim Ausfüllen des Antragsformulars für die Datentreuhandstelle unterstützen und Ihnen die Arbeit erleichtern.

Allgemeine Hinweise

- Das Formular ist mit mehreren Seiten auf den ersten Blick sehr umfangreich. Allerdings müssen nicht alle Teile immer ausgefüllt und angegeben werden. Möchten Sie beispielsweise nur vorhandene Daten pseudonymisieren lassen, füllen Sie auch nur die Formulareile unter dem Abschnitt „Pseudonymisierung“ aus.
- Geben Sie bitte immer an, ob es sich um einen Erst- oder einen Folgeantrag handelt. Die Datentreuhandstelle legt bei einem Erstantrag für ein Projekt interne Daten an und vergibt ein sogenanntes Salt, damit die pseudonymisierten Daten auch sicher sind. Und nur Daten, die zum selben Projekt gehören, lassen sich miteinander verknüpfen.

Antragstellende Person und Projektinformationen

- Tragen Sie immer Ihren vollständigen Namen und Ihre vollständigen Kontaktinformationen ein.
- Tragen Sie unter „Projektinformationen“ einen Namen für das Projekt ein, in dem die Daten genutzt werden sollen. Mit diesem Projektnamen kann die Datentreuhandstelle die DTS-internen Informationen zuordnen und das richtige Salt für die Pseudonymisierung der Daten nutzen.
- Bitte lesen Sie sich die anzugebenden Verpflichtungserklärungen aufmerksam durch. Allen Erklärungen muss zugestimmt werden, ansonsten kann die DTS die Daten nicht verarbeiten.

Kursumgebungen

- Die meisten zu pseudonymisierenden Daten beziehen sich auf eine oder mehrere Kursumgebungen auf der Moodle-Instanz „moodle.fernuni-hagen.de“. Daher listen Sie hier bitte die Kursumgebungen auf, auf welche sich die Daten beziehen.
- Sie können bis zu fünf Kursumgebungen angeben. Die Ziffern 1 bis 5 sind für die Zuordnung der weiteren Daten relevant.
- Geben Sie immer den Namen der Kursumgebung sowie die ID-Nummer der Kursumgebung an. Die ID-Nummer der Kursumgebung können Sie dem Adressfeld des Browser entnehmen, wenn Sie die Kursumgebung aufrufen. Dort steht dann beispielsweise: <https://moodle.fernuni-hagen.de/course/view.php?id=1234>
In diesem Fall ist die ID-Nummer der Kursumgebung 1234.
- Tragen Sie auch immer die modilverantwortliche Person oder die Kursbetreuer*in ein. Sind nicht Sie für die Kursumgebung verantwortlich, lassen Sie die Person digital signieren und bestätigen, dass die Daten aus der Kursumgebung genutzt werden dürfen.
- Alle an den Kursumgebungen beteiligten Beschäftigten der FernUni müssen über die Datenentnahme informiert sein und dieser zugestimmt haben. Andernfalls kann die DTS die Daten nicht verarbeiten.

Studierendenlisten, Logdaten

- Die Ziffer in der Spalte „Kursumgebung“ bezieht sich auf die im Abschnitt „Kursumgebungen“ aufgelisteten Kursumgebungen. Möchten Sie also für die Kursumgebung 1 die Logdaten haben

und für Kursumgebung 2 die Logdaten sowie die Studierendenliste, kreuzen Sie in der passenden Zeile die entsprechenden Einträge an.

- Zusätzlich können Sie für die Logdaten auswählen, in welchem Format diese bereitgestellt werden sollen.
- Die Spalte „Erledigt“ dient für interne Zwecke und soll nicht ausgefüllt werden.
- Unter „Ergänzende Angaben“ können Sie beispielsweise weitere Logdatenformate für einzelne Kursumgebungen anfragen.
- Studierendenlisten enthalten Angaben zu allen Personen, die in einer Kursumgebung mit der „Studierenden“-Rolle bzw. davon abgeleiteten Rollen eingeschrieben sind. Enthalten sind eine pseudonymisierte Nutzenden-ID, die pseudonymisierten Namen aller Gruppen, in welche die Studierenden eingetragen sind, sowie die Gruppierungen (nicht pseudonymisiert).
- Die Logdaten enthalten folgende Einträge:
 - id: fortlaufende Zählung des Events
 - hashuser/relatedhashuser: pseudonymisierte Nutzenden-ID der Person, welche das Event ausgelöst hat
 - courseid: Kurs-ID in Moodle
 - timecreated: Unix-Timestamp
 - eventname: Kombination aus component, target und action
 - component: Moodle-Komponente, auf die sich das Event bezieht, bspw. Kursumgebung oder Forum
 - action: komponentenspezifische Aktion
 - target: Zielobjekt der Aktion
 - objectid: ID des Zielobjekts
 - crud: create – read – update – delete
 - edulevel: 0 (andere), 1 (teaching), 2 (participating)

Forendaten

- Die Ziffer in der Spalte „Kursumgebung“ bezieht sich auf die im Abschnitt „Kursumgebungen“ aufgelisteten Kursumgebungen. Möchten Sie also für die Kursumgebungen 1 und 3 die Forendaten erhalten, setzen Sie für die entsprechenden Zeilen ein Kreuz in der Spalte „Forendaten“.
- In der Spalte „Auf diese Foren beschränken“ können Sie die Namen der Foren eintragen, für die Sie die Forendaten anfragen.
- Die Forendaten enthalten die Forenbeiträge der Nutzenden, die in der entsprechenden Kursumgebung mit der „Studierenden“-Rolle eingetragen sind.
- Die Forendaten enthalten folgende Einträge:
 - id: Identifikationsnummer des Forenbeitrags
 - hashuser: pseudonymisierte Nutzenden-ID der Person, die den Forenbeitrag verfasst hat
 - discussion: Identifikationsnummer des Diskussionsstrangs
 - parent_id: ID übergeordneten Forenbeitrags
 - created: Unix-Timestamp
 - subject: Beitragstitel
 - message: pseudonymisierter Beitragstext
 - forum_id: ID des Forums
 - Forumname: Name des Forums
- Die Forenbeiträge werden pseudonymisiert und sog. personenbezogene Daten sowie Personally Identifiable Information (PII) aus dem Text entfernt. Das geschieht in mehreren Stufen:

- Zunächst wird intern eine Liste mit den Namen aller Personen erstellt, die in der entsprechenden Kursumgebung mit der „Studierenden“-Rolle eingetragen sind. Dabei werden automatisch verschiedene Varianten erstellt: Vor- und Nachname zusammen, nur Vorname, nur Nachname, Schreibungen mit Genitiv-S. Diese Einträge werden für eine lexikalische Filterung des Beitragstextes genutzt. Alle Treffer werden gegen den Platzhalter „<PERSON>“ ausgetauscht. Aus „Liebe Grüße, Peter“ wird dabei „Liebe Grüße, <PERSON>“.
- Häufig genutzte Spitznamen und andere Namensvarianten werden mittels einer internen Liste ausgefiltert und ebenfalls gegen „<PERSON>“ ausgetauscht.
- Danach werden mit regulären Ausdrücken Mailadressen, Telefon- und Handynummern sowie URLs herausgefiltert und gegen die Platzhalter „<EMAIL>“, „<PHONE>“ und „<URL>“ ausgetauscht.
- Im letzten Schritt durchlaufen alle Forenbeiträge eine lokale Instanz von Microsoft Presidio (<https://microsoft.github.io/presidio/>), die ebenfalls Namen, Mailadressen, Telefon- und Handynummern sowie URLs ausfiltert. Presidio ist eine Open-Source-Software von Microsoft, die mit verschiedenen Ansätzen PII in einem Text identifiziert und gegen die o.g. Platzhalter austauscht. Presidio wurden dazu an die Bedarfe der Datentreuhandstelle angepasst und läuft in einem Docker-Container auf dem Rechner der Datentreuhandstelle.

Zuordnen und Pseudonymisieren

- Sie können andere Daten wie Befragungsdaten oder Berichte aus Moodle pseudonymisieren lassen. Diese Daten müssen Sie selbst beschaffen der Datentreuhandstelle als Datei übergeben, wobei bestimmte Dinge zu beachten sind.
- Jede Datei muss als Zuordnungsvariable entweder die Moodle-ID oder E-Mail-Adresse enthalten. Nur so ist es möglich, einen Eintrag in der Datei passend zu pseudonymisieren.
- Jeder Datei muss dazu eine der bis zu fünf angegebenen Kursumgebungen zugeordnet werden. Die Datentreuhandstelle versucht dann, jeden Eintrag in der Datei einer Person in der angegebenen Kursumgebung zuzuordnen und tauscht die Zuordnungsvariable gegen die pseudonymisierte Nutzenden-ID aus. Somit können Befragungsdaten mit Moodle-Nutzungsdaten datenschutzkonform verknüpft werden, ohne dass die Daten einer natürlichen Person zugeordnet werden können.
- Als „Dateiname“ tragen Sie den Namen der Datei ein, die durch die Datentreuhandstelle verarbeitet werden soll.
- Im Feld „Bezieht sich auf die Kursumgebung“ tragen Sie bitte die Ziffer 1 bis 5 für einen der Einträge im Abschnitt „Kursumgebungen“ ein.
- Im Feld „Zuordnungsvariable“ geben Sie den Namen der Variable in der Datei ein, welche die Zuordnungsvariable ist und entweder die Moodle-ID oder E-Mail-Adresse enthält.
- In der Spalte „Zuordnungsvariable ist“ geben Sie an, ob die Zuordnungsvariable die Moodle-ID oder E-Mail-Adresse enthält.
- Zusätzlich können Sie im Feld „Zu löschende Variable(n)“ die Namen von weiteren Variablen angeben, die durch die Datentreuhandstelle gelöscht werden sollen.

Pseudonymisierung

- In wenigen Fällen haben Sie Daten, in denen die Matrikelnummer von Studierenden enthalten sind und die sie mit anderen Daten verknüpfen möchten. Das kann bspw. eine Notenliste sein. In diesen Fällen können Sie die Datei durch die Datentreuhandstelle pseudonymisieren lassen, um bspw. Prüfungsnoten zusammen mit Moodle-Logdaten zu analysieren.

- Die Spalte „Nr“ enthält einfach eine fortlaufende Zählung.
- Im Feld „Datei“ geben Sie den genauen Namen der zu verarbeitenden Datei und im Feld daneben den genauen Namen der zu pseudonymisierenden Variable/n ein.

Moodle-Berichte

Das Lernmanagementsystem Moodle stellt Lehrenden verschiedene Berichte und andere Daten zur Verfügung, die Sie durch die Datentreuhandstelle verarbeiten lassen können.

Bewertungsübersicht

- Bewertungsübersicht enthalten Punkte oder andere Bewertungen, die Studierende bei der Bearbeitung verschiedener Aktivitäten erhalten. Diese Bewertungsübersicht können Lehrende herunterladen und durch die Datentreuhandstelle pseudonymisieren lassen.
- Rufen Sie dazu die gewünschte Kursumgebung auf und wählen Sie in der Kursnavigation den Eintrag „Bewertungen“.
- Aus der Auswahlliste „Bewertungsübersicht“ wählen Sie anschließend den Eintrag „Export“ aus.
- Als Format wählen Sie unter „Exportieren als“ entweder „Excel-Datei“ oder „Textdatei“ aus.
- Bestätigen Sie die Datenschutzbestimmungen und wählen anschließend „Herunterladen“ in der Fußzeile aus.
- Die zu pseudonymisierende Variable ist: . Geben Sie diese im Antragsformular an.

Quiz-Ergebnisse

- Die Ergebnisse von einzelnen Quizzes können Sie ebenfalls exportieren und pseudonymisieren lassen.
- Rufen Sie das gewünschte Quiz auf und wählen Sie dann den Eintrag „Ergebnisse“.
- Sie können die Berichtstabelle direkt herunterladen oder anpassen. Sie können sich bspw. alle Versuche oder nur den besten Versuch pro Student*in anzeigen lassen.
- Laden Sie anschließend die Tabellendaten entweder als „Microsoft Excel (.xlsx)“ oder als „Komma separierte Werte (.csv)“ herunter.
- Die zu pseudonymisierende Variable heißt „ID-Nummer“. Geben Sie diese im Antragsformular an.

Aktivitätsabschluss

- Wenn in einer Kursumgebung die Abschlussverfolgung aktiviert ist, ist im Bericht „Aktivitätsabschluss“ hinterlegt, wann welche Aktivität erfolgreich abgeschlossen wurde.
- Rufen Sie dazu von der Kursstartseite den Eintrag „Bericht“ > „Aktivitätsabschluss“ auf und laden Sie den Bericht „Als Excel-Datei (.csv)“ herunter.
- Die zu pseudonymisierende Variable heißt „ID-Nummer“. Geben Sie diese im Antragsformular an.
- Außerdem sollten die Variablen „Vor-/Nachname“ und „E-Mail-Adresse“ entfernt werden.

Befragungen in Moodle

- Befragungen können in Moodle mit der Questionnaire- bzw. Befragungs-Aktivität durchgeführt werden.
- Achten Sie darauf, dass die Befragung nicht anonym durchgeführt wird, wenn die Daten später mit anderen Daten verknüpft werden sollen. Überprüfen Sie dazu in den Einstellungen > Antwortoptionen > Abgabety, ob dort die Option „Vollständiger Name“ ausgewählt ist und korrigieren Sie die Einstellungen falls erforderlich.

- Um nach dem Abschluss der Befragung die Antworten herunterzuladen, rufen Sie auf der Befragungsseite oder über die Kontextnavigation die Option „Alle Antworten anzeigen“ auf.
- Wählen Sie dann „Herunterladen“ aus, entfernen die Häkchen bei „Auswahlcode einbeziehen“ und „Auswahltext einbeziehen“ aus und laden die Daten entweder als CSV oder als XLSX herunter.
- Die zu pseudonymisierende Variable heißt „Benutzername“, außerdem sollten die Variablen „Institution“, „Abteilung“, „Lernumgebung“, „Nutzer-ID“ sowie „Vollständiger Name“ entfernt werden.