

IDZ-Skript zum Auslesen aggregierter Daten aus einer digitalen fallbasierten Dokumentation

Stand 30. September 2023

Hintergrund des Skriptes

Das Skript wurde vom Institut für Demokratie und Zivilgesellschaft ([IDZ Jena](#)) für die Studie „Durchführung einer Abfrage unter Antidiskriminierungsberatungsstellen zum Beratungsaufkommen und zu Erkenntnissen aus der Beratungstätigkeit“ entwickelt. Auftraggeberin dieser Studie ist die Antidiskriminierungsstelle des Bundes.

Weitere Informationen zu der Studie finden Sie [hier](#).

Aufgabe des Skriptes

Das Skript wurde entwickelt, um zwei Funktionen zu erfüllen:

Synchronisierung der Dokumentationskategorien der Beratungsstelle mit den Kategorien der Erhebung

Unterschiedliche Beratungsstellen sprechen unterschiedliche Dokumentations“sprachen“: Ein und derselbe Lebensbereich kann „Arbeit“, „Arbeit und Beschäftigung“ oder „Arbeitsverhältnisse“ heißen.

Die Software leitet Beratungsstellen in einem Prozess an, in dem Sie Ihre Kategorien (z.B. Arbeit und Beschäftigung) den Kategorien der Erhebung (Arbeit und berufliche Ausbildung) zuordnen.

Die Software lernt auf diese Weise, wo in einem Datensatz welche Informationen zu finden sind und übersetzt zwischen der Kategoriensprache der Beratungsstelle und der Kategoriensprache der Erhebung.

Aufbereitung und Zusammenfassung der für die Erhebung relevanten Informationen aus einem Dokumentationsdatensatz

Anschließend fasst das Skript die für die Erhebung relevanten Daten aus dem Datensatz der Beratungsstelle in Häufigkeitstabellen zusammen (Datenaggregation). Zum Beispiel „zählt“ es wie häufig Sie 2022 Beratungsfälle im Bereich Arbeit und berufliche Ausbildung hatten und stellt diese Ergebnisse als Tabellen zusammen.

Diese Tabellen werden als Excel-Dateien lokal (auf dem Rechner) gespeichert.

Zusätzlich arbeitet das Skript diese Informationen grafisch, beispielsweise als Häufigkeitsverteilungen, auf. Diese Dateien geben der Beratungsstelle einen Überblick und sind als Hilfe für deren Arbeit gedacht. Auch diese Dateien werden lokal (auf dem Rechner) gespeichert.

Grundsätzliche Informationen zum Skriptes

- ➔ Das Skript wurde eigens für die Erhebung vom IDZ Jena entwickelt und ist in Python programmiert. Es ist ein eigenständiges Programm, das ohne zusätzliche Installation auf allen Rechnern mit einem Windows-Betriebssystem genutzt werden kann.
- ➔ Das Skript wird über den Server des IDZ Jena sicherheitsverschlüsselt zum Download angeboten. Hierfür ist ein Internetzugang nötig. Link: https://www.idz-jena.de/fileadmin/media/Skript_Erhebung.exe
- ➔ Die Nutzung des Skriptes erfolgt lokal auf dem jeweiligen Rechner. Dafür ist kein Internet notwendig. Die Speicherung der Ergebnisdateien erfolgt lokal. Es werden keine Daten automatisch weitergeleitet.
- ➔ Die Software wurde IDZ-intern intensiv auf Ihre Funktionsfähigkeit und Sicherheit geprüft. Eine externe Zertifizierung durch Windows oder einen unabhängigen Anbieter konnte im Rahmen des Projektes nicht umgesetzt werden.
- ➔ Der Code des Skriptes ist auf Github unter <https://github.com/dabaIDZ/Skript-Erhebung-Dokudaten> abgelegt und kann dort jederzeit eingesehen, geprüft und kommentiert werden.

Vorgehensweise bei der Nutzung des Skriptes

➔ Schritt 1: Download des Skriptes von IDZ-Server

Die nutzende Organisation lädt das Skript vom Server des IDZ Jena.

➔ Schritt 2: Start des Skriptes auf dem organisationseigenen Rechner

Das Skript kann als eigenständiges Programm auf einer Windows-Oberfläche gestartet werden. Die gesamte Anwendung des Skriptes geschieht lokal, auf dem Rechner. Für die Nutzung ist kein Internet- oder Browserzugang nötig.

Die Informationen der Beratungsstelle (Dokumentationsdaten) werden zu keiner Zeit während der Anwendung des Skriptes oder im Anschluss an Dritte (z.B. das IDZ) weitergeleitet. Sämtliche Verarbeitungs- und Speicherungsprozesse laufen ausschließlich auf dem Rechner und in der Kontrolle der Organisation. Ein Zugriff von Außen ist nicht möglich. Auch das IDZ hat zu keiner Zeit und in keiner Form Zugriff auf die fallbezogenen Ausgangsdaten, Verarbeitungs- und Speicherprozesse.

➔ Schritt 3: Anwendung des Skriptes

Auf einer graphischen Nutzer*innen-Oberfläche wird die Organisation Schritt für Schritt durch die Anwendung geleitet:

- (1) Laden der Datei mit den fallbezogenen Dokumentationsdaten: Die lokal auf dem Rechner befindliche Datei (.xls oder .csv-Tabelle) mit allen Rohdaten wird aufgerufen und dem Skript der Zugriff auf die darin gespeicherten Daten ermöglicht.
- (2) Auswahl der Fälle für das Jahr 2022: In einem ersten Schritt wird die Organisation gebeten, die Fälle auszuwählen, die in den Erhebungszeitraum fallen. Nur diese werden anschließend genutzt. Die Rohdaten/ Quelldatei muss somit nicht im Vorfeld auf Fälle aus dem Jahr 2022 begrenzt werden.

- (3) Zuweisung der Kategorien: Die Organisation wird systematisch durch die für die Erhebung relevanten Kategorien geführt und wählt jeweils manuell die korrespondierende Kategorie (in der Regel eine Tabellenspalte) aus Ihrer Dokumentationsdatei. Z.B.: „Arbeit“...entspricht der eigenen Kategorie „Arbeit und Beschäftigung“.
Auf diese Weise „erfährt“ das Skript wo es in der Ausgangstabelle spezifische Daten findet, die für den nächsten Schritt der Auswertung benötigt werden.
Für Nutzer_innen der advd-Software entfällt dieser Schritt, weil die Zuordnung der Kategorien im Vorfeld bereits durch das IDZ erfolgte und voreingestellt ist (Sonderform des Skriptes).
- (4) Datenverarbeitung: Nachdem die Kategorien der Erhebung und die Kategorien des jeweiligen Dokumentationssystems miteinander verknüpft sind, erfolgt die (lokale) Datenverarbeitung: Die für die Erhebung relevanten aggregierten Daten werden aus der Quelldatei ausgelesen. Z.B. Wie viele Fälle im Bereich Arbeit, Bildung, ...? Wie viele Fälle rassistischer Diskriminierung im Bereich Arbeit? Wie viele Fälle von Diskriminierung im Kontext von Behinderung im Bereich Arbeit usw.
Es werden a) nur die Daten verarbeitet, die für die konkreten Fragestellungen wichtig sind und b) die Ausgabe umfasst nur aggregierte (zusammengefasste) Zahlen von denen nicht auf Einzelfälle zurückgeschlossen werden kann.
- (5) Ausgabe der Ergebnisse: Die Ergebnisse werden auf zweierlei Art ausgegeben und gespeichert: (A) in Form eines Berichtes (pdf) für die Beratungsstellen. Diese Datei enthält zentrale Ergebnisse in einer Übersicht für die Nutzung durch die Beratungsstelle und verbleibt bei der Beratungsstelle. Und (B) eine Reihe von .csv-Dateien (Datentabellen), die Erhebung im engeren Sinne, die als .zip-Datei an das IDZ geschickt werden soll.

➔ **Schritt 4: Bericht für die Nutzung durch Beratungsstelle, .zip-Datei für IDZ**

Im Rahmen der Auswertung erstellt das Skript eine Reihe von Ergebnisdiagrammen (PNG-Dateien), die von der Beratungsstelle für ihre Arbeit genutzt werden kann sowie eine Reihe von Excel-Dateien, die die aggregierten Daten für die Erhebung enthalten. Diese Excel-Dateien werden zusätzlich noch einmal als Zip-Datei ausgegeben. Alle Dateien werden lokal auf dem Rechner an einer von den Nutzer*innen definierten Ablageort wie folgt gespeichert.

Ordner: 01 Informationen für die Beratungsstelle (enthält alle Ergebnisdiagramme)

Ordner: 02 Daten für die Erhebung

Zip-Datei: An IDZ senden

Die .zip-Datei wird für die Erhebung benötigt und soll von der Beratungsstelle anschließend per Email an das IDZ weiterleitet werden.

Erst mit dieser Aktion, die bewusst von der Organisation initiiert werden muss, verlassen Daten den eigenen Rechner und werden für das IDZ zugänglich. Der Umfang und der Inhalt dieser Daten ist streng begrenzt und transparent. Es handelt sich ausschließlich um aggregierte Daten, die weder personenbezogene Daten enthalten noch eine Rekonstruktion von konkreten Einzelfällen erlauben.

Transparenz und Kontrolle

Die Transparenz wird sichergestellt durch:

- die Veröffentlichung des Quellcodes des Skriptes parallel zum Skript (kann durch fachkundige Personen nachvollzogen werden)
- eine Übersicht aller Datentabellen, die in der Zip-Datei an das IDZ enthalten sind

- die Möglichkeit, die Zip-Datei zu entpacken und jede einzelne Datentabelle auf ihren Inhalt hin zu überprüfen und ggf. auch einzelne Datentabellen zu löschen, die nicht an das IDZ weitergegeben werden sollen

Datenschutz

Die Verarbeitung der fallbezogenen Rohdaten erfolgt vollständig in der Kontrolle der Beratungsstelle. Das IDZ hat in keiner Form Zugriff auf diese Daten.

Die als Ergebnis gespeicherten und an das IDZ weitergeleiteten Daten sind ausschließlich aggregiert. Eine Rekonstruktion von Einzelfällen ist nicht möglich.

Die Abfrage umfasst keine personenbezogene Daten. Eine Verarbeitung und gar Speicherung personenbezogener Daten, die in den Rohdaten enthalten sind, ist durch die Struktur des Skriptes ausgeschlossen.

Zuordnung von Datensätzen zu Beratungsstellen

Die Verwendung des Skriptes ist eingebettet in die grundsätzliche Erhebung (die zusätzlich noch aus einem Online-Fragebogenteil besteht). Für die Erhebung unterscheiden wir bezogen auf die Zuordnung von Datensätzen zu Beratungsstellen zwei Phasen: Im Rahmen der Datenerhebung und -analyse werden die aggregierten Daten aus dem Skript und der Onlinebefragung den jeweiligen Beratungsstellen zugeordnet sein. Neben dem IDZ wird auch die Antidiskriminierungsstelle des Bundes als Auftraggeberin der Abfrage Zugang zu diesen Daten erhalten. Bei der Veröffentlichung der Ergebnisse der Abfrage wird sichergestellt werden, dass die Zuordnung von Daten zu einzelnen, spezifischen Beratungsstellen weder direkt noch indirekt möglich ist.

Fragen

Für technische und organisatorische Fragen zum Skript stehen wir Ihnen gern zur Verfügung. Sie erreichen das Studien-Team unter:

E-Mail: studie-dokudaten@idz-jena.de

Telefon: 0157 - 537 132 67

Projektleitung: Daniel Bartel