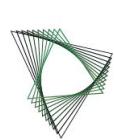


Modul 1: Einführung, Problemerfassung und Daten verstehen

Angewandte Datenanalyse für die öffentliche Verwaltung in Bayern (ADA Bayern)
www.ada-oeffentliche-verwaltung.de



BERD
@NFDI



Bayerisches Staatsministerium
für Digitales



Über uns



Prof. Dr. Frauke Kreuter



Dr. Malte Schierholz



Dr. Marcel Neunhoeffer



Dr. Heidi Seibold

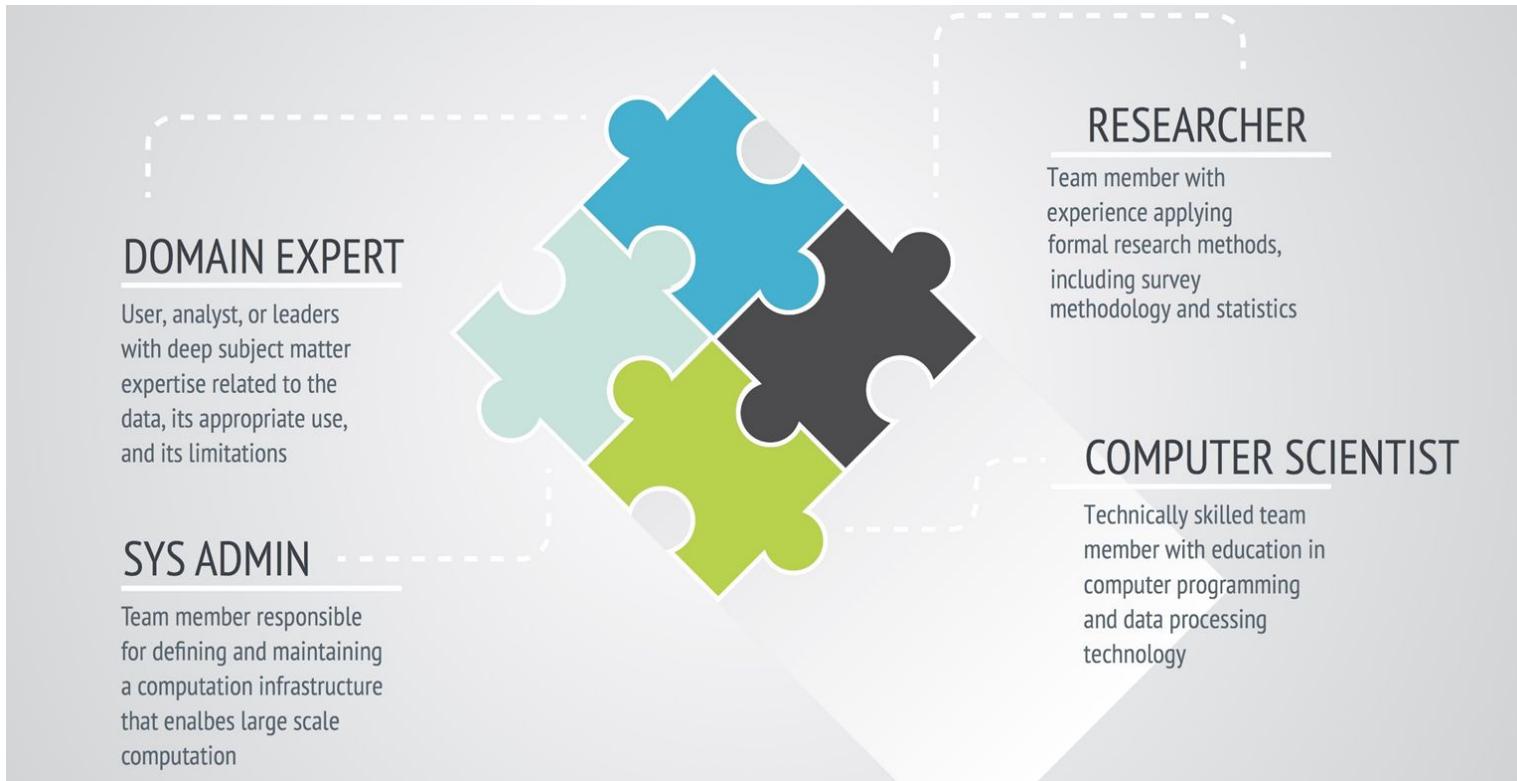


Felix Henninger

Über euch/Sie

- Name
- Institution
- Erfahrung mit Daten
- Fun Fact!

Data Science is a Team Sport!



Teamarbeit

Die Teams arbeiten an einem Projekt, das am Ende in einer Präsentation vorgestellt werden soll:



3 Slides pro Team:

- Ergebnisse
- Plan für Umsetzung
- Zukunftsvision: Was bräuchte man noch für das "perfekte" Archivierungs-System?

Die Teams

Team 1 - Raum Nightingale

Dominique Mergen, Hannah Hien, Michael Puchta, Markus Schmalzl, Heidi Seibold

Team 2 - Raum Singer

Christian Pfrang, Andreas Nestl, Andreas Hutterer, Jens Ulken, Frauke Kreuter

Team 3 - Raum Gebru

Daniel Gietl, Michael Unger, Johannes Stoiber, Tilman Janzarik, Malte Schierholz

Team 4 - Zoom

Christian Brück, Claudia Kalesse, Verena Christl, Sascha Bauer, Franziska Armbruster, Paul Manthey, Felix Henninger

Team 5 - Zoom

Gunther Friedrich, Ulrike Claudia Hofmann, Rhea Nguyen, Anita Bott, Dominik Mayer, Carolin Stimmelmayr, Marcel Neunhoeffer

Evidenzbasierte Verwaltung



THE PROMISE OF EVIDENCE-BASED POLICYMAKING

Report of the Commission on Evidence-Based Policymaking



EXECUTIVE OFFICE OF THE PRESIDENT
OFFICE OF MANAGEMENT AND BUDGET
WASHINGTON, D.C. 20503

June 4, 2019

M-19-18

MEMORANDUM FOR THE HEADS OF EXECUTIVE DEPARTMENTS AND AGENCIES

FROM: Russell T. Vought
Acting Director

SUBJECT: Federal Data Strategy - A Framework for Consistency

AUTHENTICATED

- *Building a Culture that Values Data and Promotes Public Use (practices 1-10)*
- *Governing, Managing, and Protecting Data (practices 11-26)*
- *Promoting Efficient and Appropriate Data Use (practices 27-39)*

Public Law 115–435
115th Congress

An Act

To amend titles 5 and 44, United States Code, to require Federal evaluation activities, improve Federal data management, and for other purposes.

Jan. 14, 2019
[H.R. 4174]

Be it enacted by the Senate and House of Representatives of the United States of America in Congress assembled,

SECTION 1. SHORT TITLE; TABLE OF CONTENTS.

- (a) **SHORT TITLE.**—This Act may be cited as the “Foundations for Evidence-Based Policymaking Act of 2018”.
(b) **TABLE OF CONTENTS.**—The table of contents for this Act

Foundations for
Evidence-Based
Policymaking Act
of 2018.
5 USC 101 note.

[ENGLISH](#)[FRANÇAIS](#)[KONTAKT](#)[DATENSCHUTZHINWEIS](#)

...

[☰ Menü](#) | [Bundesregierung](#) | [Startseite](#) [Suche](#)

Datenstrategie der Bundesregierung

Eine Innovationsstrategie für gesellschaftlichen Fortschritt und nachhaltiges Wachstum

- Kabinettfassung, 27. Januar 2021



...



Broschüre, 120 Seiten

Stand: 27. Januar 2021

Sprachen: Deutsch

Bundesministerium für Umwelt, Naturschutz, nukleare Sicherheit und Verbraucherschutz 2022

Datenlabor

Das BMUV initiiert und betreut unter anderem zwei zentrale Bausteine der Datenstrategie der Bundesregierung, die zugleich zentrale Modernisierungsvorhaben des BMUVs sind: das „BMUV Datenlabor“ und das „Anwendungslabor für KI und Big Data (KI-Lab)“.

Ziel beider Vorhaben ist, die Datengrundlage für die politische Entscheidungsfindung des Ministeriums zu verbessern. Beide Maßnahmen sind thematisch eng miteinander verbunden und sollen in Zukunft noch besser voneinander profitieren.

Ziel des Projekts war daher die bedarfs- und nutzerzentrierte Entwicklung von gemeinsamen Prozessen und Dienstleistungen beider Labore mit Fokus auf den Aufbau des Datenlabors.



24.05.2022

Hintergrundinformation



Was ist eigentlich die Frage? Problemstellung definieren.



Ziel: Definiere das Projektziel



Aktion: Welche Maßnahmen/Handlungen wird dieses Projekt informieren?



Daten: Welche Daten stehen dazu intern zur Verfügung? Welche müssen extern ergänzt werden? Was brauchst du um an die Daten zu kommen?



Analyse: Welche Analyse ist notwendig? Handelt es sich dabei um eine Beschreibung, Intervention, Vorhersage oder Verhaltensänderung

Ethische Aspekte: Privacy, Transparenz, Gleichstellung, Verantwortlichkeiten

Es beginnt mit einer Frage...

Wir haben 100 KiTa Plätze. Wer soll die bekommen?

Welche Flächen müssen wir in Zukunft wegen Hochwasser anders bewirtschaften

Wir haben 100 Professuren in der High-Tech Agenda finanziert. War das sinnvoll?

Wie verteilen wir die Impfdosen?

Es gibt Berichte von Sozialarbeitern. Wo müssen wir intervenieren?

Sollen wir diese Maßnahme nochmal machen?

Was soll archiviert werden?

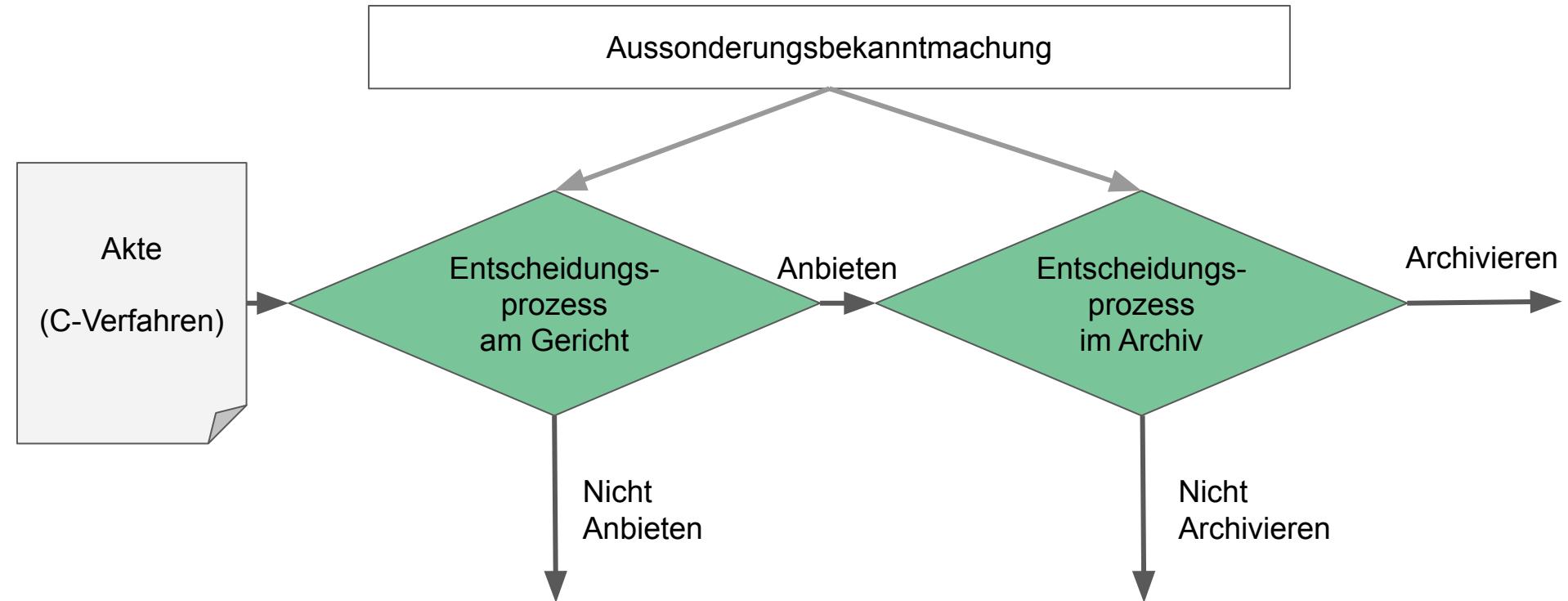
Was ist eigentlich die Frage? Problemstellung definieren.

Hier:



Entscheidungsprozess heute

Wie sieht der Prozess für Rechtspfleger:innen aus? Wie für Archivar:innen?



Fragen beantworten mit Daten



- In den US verbleiben weniger als die Hälfte der HIV-Positiven in Versorgungsprogrammen.
- KI hilft ambulanten vorherzusagen, wer aus Vorsorge herausfällt.
- Ressourcen für Interventionen werden entsprechend des Ausfallrisikowertes geplant

Ramachandran, A., Kumar, A., Koenig, H. et al. Predictive Analytics for Retention in Care in an Urban HIV Clinic. *Sci Rep* 10, 6421 (2020).
<https://doi.org/10.1038/s41598-020-62729-x>

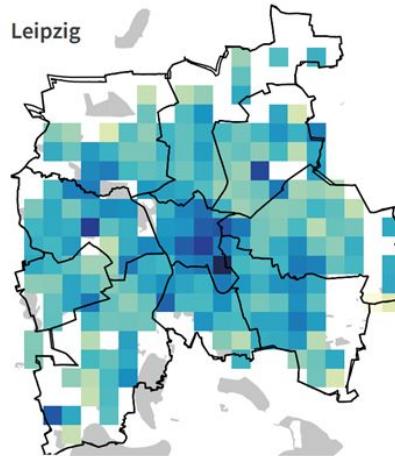
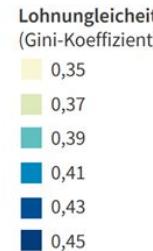
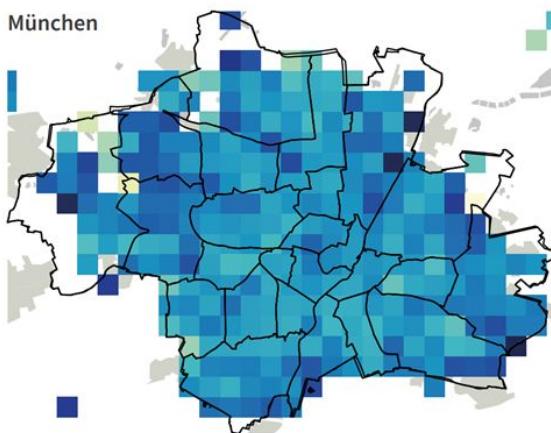
- Kooperation: Johnson County, Kansas Carnegie Mellon University
- Ziel: Spirale unbehandelter psychiatrischer Erkrankungen und Verhaftungen zu durchbrechen.
- KI-gestützte Vorhersage der Wahrscheinlichkeit ins Gefängnis zu kommen
- Personen mit hohem Risikowert werden je nach vorhandenen Ressourcen priorisiert behandelt



Rodolfa, K.T., Lamba, H. & Ghani, R. Empirical observation of negligible fairness–accuracy trade-offs in machine learning for public policy. *Nat Mach Intell* 3, 896–904 (2021).
<https://doi.org/10.1038/s42256-021-00396-x>

Fragen beantworten mit administrativen Daten

Kleinräumige Verteilung der Lohnungleichheit in München und in Leipzig 2017



■ geografische Flächen, die schwach besiedelt, bewaldet oder Gewässer sind

Die Datenbasis beruht auf den administrativen Daten der Bundesagentur für Arbeit und zeigt 1x1-Kilometer-Gitterzellen. Insgesamt werden für München 294 und für Leipzig 218 unzensierte Gitterzellen (mit mehr als 20 Personen) abgebildet. Dargestelltes Maß für die kleinräumige Lohnungleichheit ist der zellenspezifische Gini-Koeffizient.

Interpretationshilfe: Hellere Gitterzellen weisen auf einen niedrigeren Gini-Koeffizienten – also niedrigere Lohnungleichheit – hin, dunklere auf einen höheren Gini-Koeffizienten. Flächen ohne Gitterzellen sind nicht oder zu schwach besiedelt und somit censiert. Die zugrunde liegenden Karten zeigen schwach besiedelte und bewaldete Flächen sowie Gewässer (grau). Schwarze Linien kennzeichnen die einzelnen Bezirke.

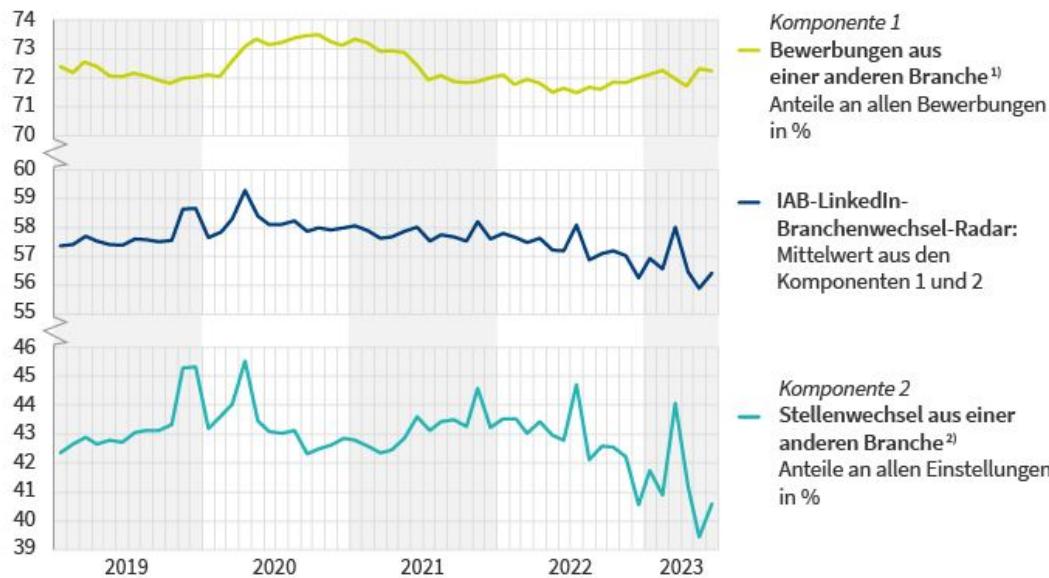
Quellenangabe: GridAB v2.1, eigene Darstellung. © IAB

“Die
Lohnungleichheit
ist in
ostdeutschen
Nachbarschaften
kleiner als in
westdeutschen”

<https://doku.iab.de/kurzber/2023/kb2023-09.pdf>

Fragen beantworten mit (den richtigen) Daten

Abb: Das IAB-LinkedIn-Branchenwechsel-Radar



¹⁾ LinkedIn-Mitglieder, die sich mit dem Online-Bewerbungsformular von LinkedIn auf eine Stelle in einer anderen Branche bewerben, als sie laut Mitgliederprofil zum Zeitpunkt der Bewerbung beschäftigt sind.

²⁾ LinkedIn-Mitglieder, in deren Mitgliederprofil vor Antritt der neuen Stelle eine Stelle in einer anderen Branche eingetragen war.

“Das IAB-LinkedIn-Branchenwechsel-Radar: „Great Resignation“ ist kein Trend”

Fragen beantworten mit kombinierten Daten

Average adult income rank
for a poor child

60th percentile

50th

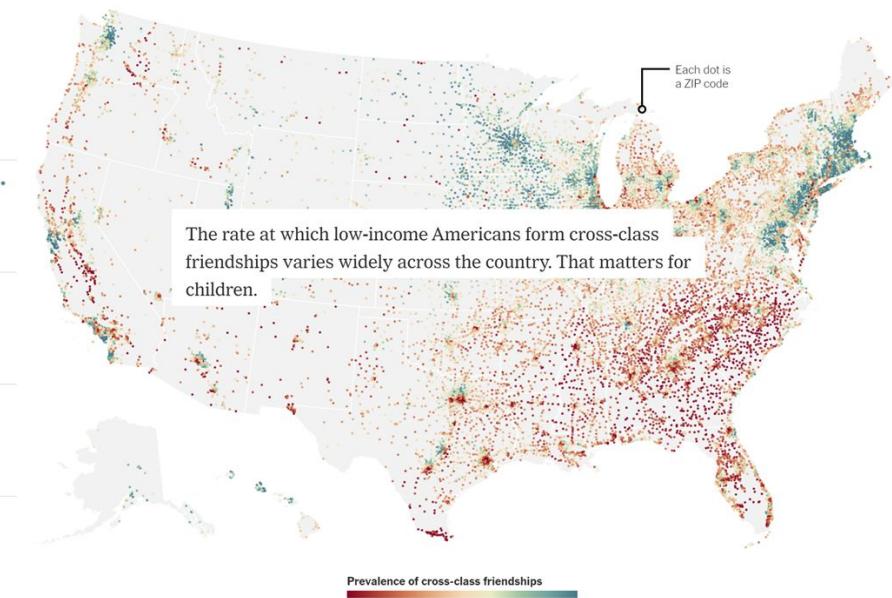
40th

30th

Winsted, Minn. (55395)
High mobility

The new research has an answer: economic connectedness. In places where poor people have more rich friends, outcomes are better.

Conley, Ga. (30288)
Low mobility



US-Steuerdaten kombiniert mit Facebook-Daten
<https://www.nature.com/articles/s41586-022-04996-4>

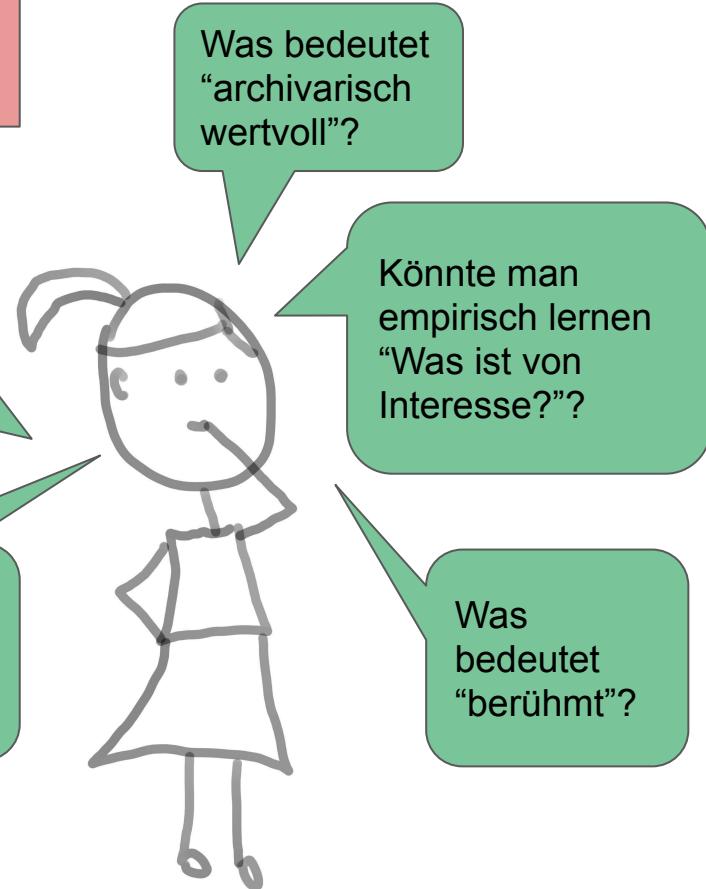
Teamarbeit 1

Expertenwissen einbinden!

Datenarbeit = Teamarbeit

1. Besprechen Sie die Expertisen der Team-Mitglieder
2. Besprechen Sie: Wie kann man entscheiden, welche Akten langfristig archiviert werden sollen?
3. Was sind die Kriterien für "archiwürdig"?

Nutzen Sie für die Diskussion auch die Gesetzlichen Vorgaben.



Die Teams

Team 1 - Raum Nightingale

Jens Ulken, Hannah Hien, Michael Puchta, Markus Schmalzl, Heidi Seibold

Team 2 - Raum Singer

Christian Pfrang, Andreas Nestl, Andreas Hutterer, Dominique Mergen, Frauke Kreuter

Team 3 - Raum Gebru

Daniel Gietl, Michael Unger, Johannes Stoiber, Tilman Janzarik, Malte Schierholz

Team 4 - Zoom

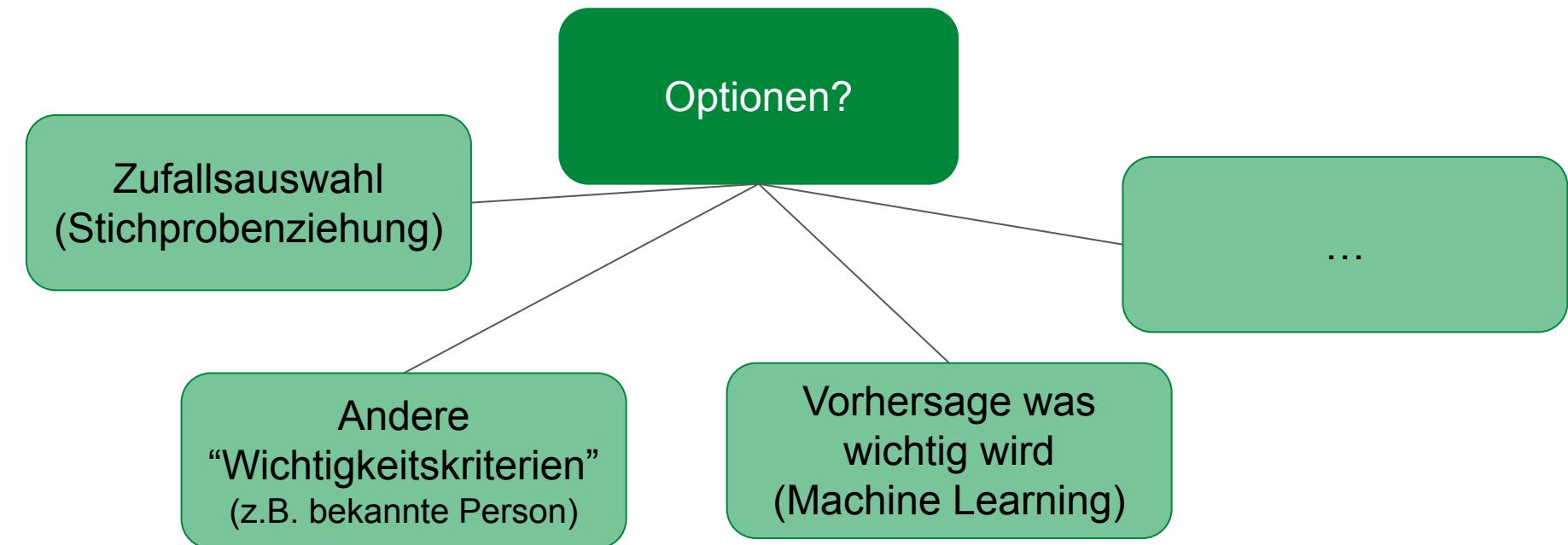
Christian Brück, Claudia Kalesse, Verena Christl, Sascha Bauer, Franziska Armbruster, Paul Manthey, Felix Henninger

Team 5 - Zoom

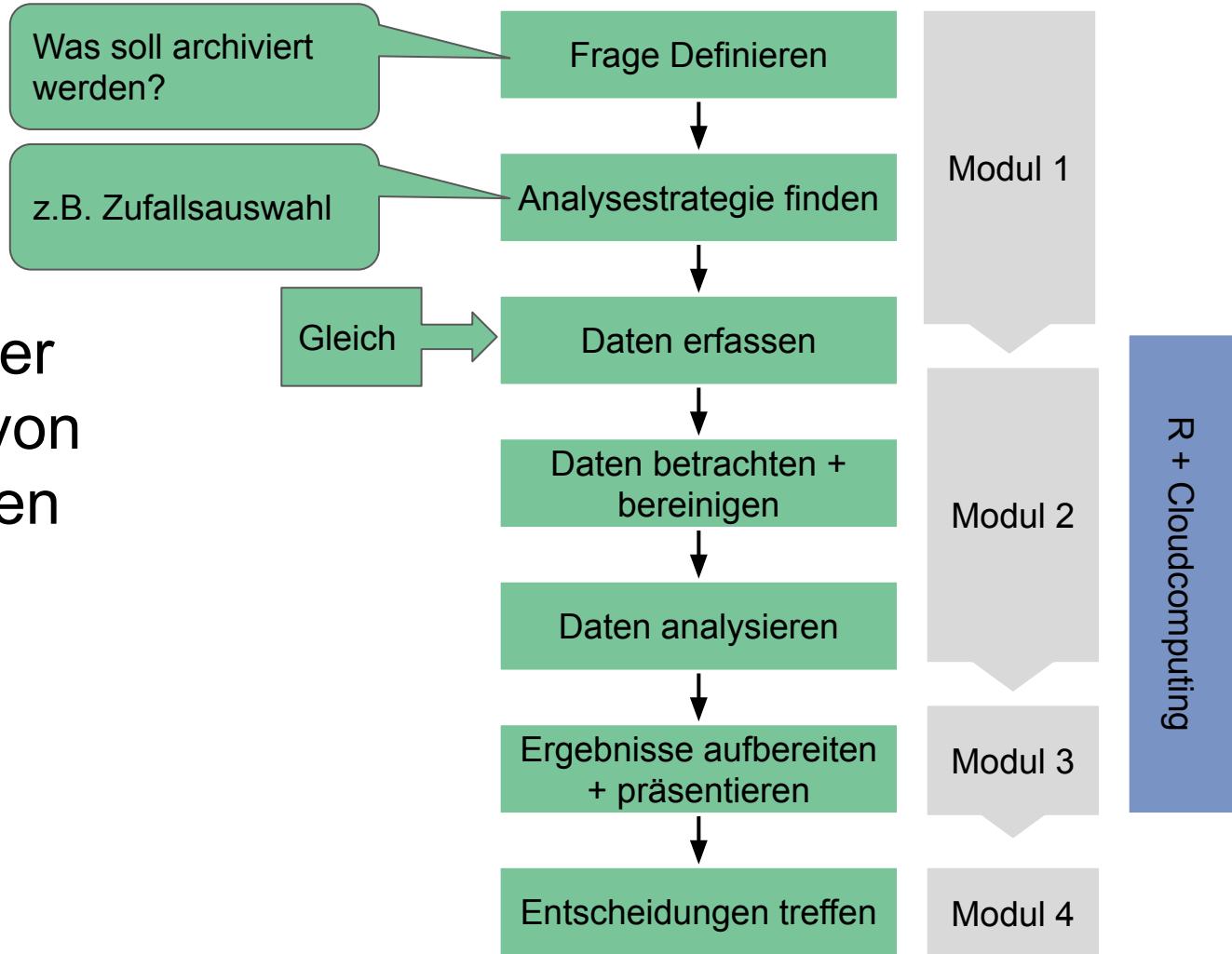
Gunther Friedrich, Ulrike Claudia Hofmann, Rhea Nguyen, Anita Bott, Dominik Mayer, Carolin Stimmelmayr, Marcel Neunhoeffer

Daten nutzen, um eigene Probleme zu lösen.

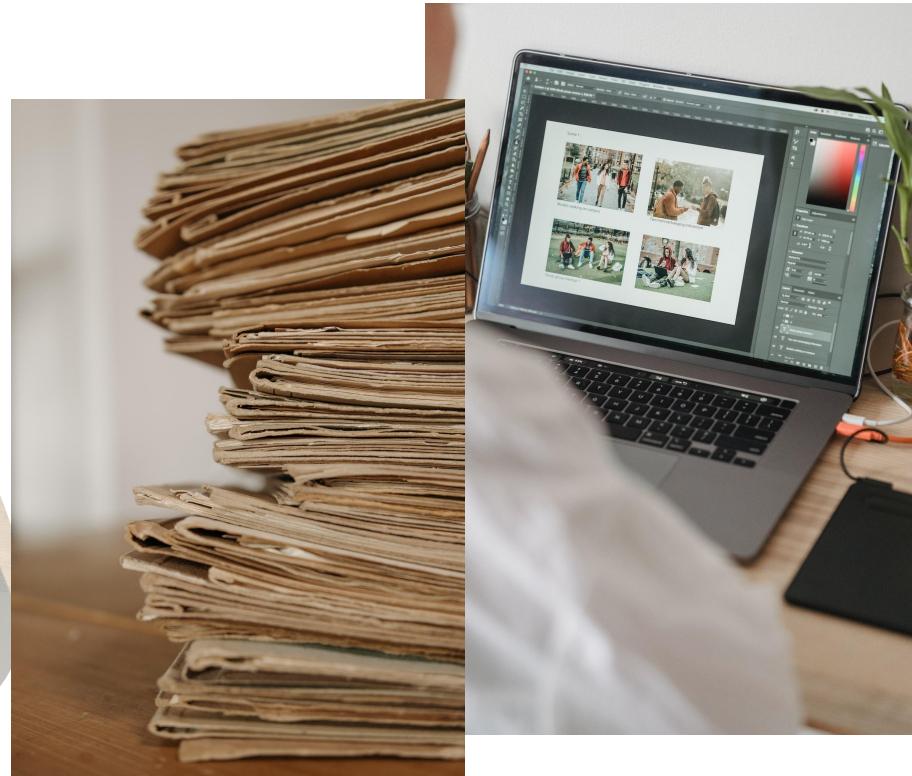
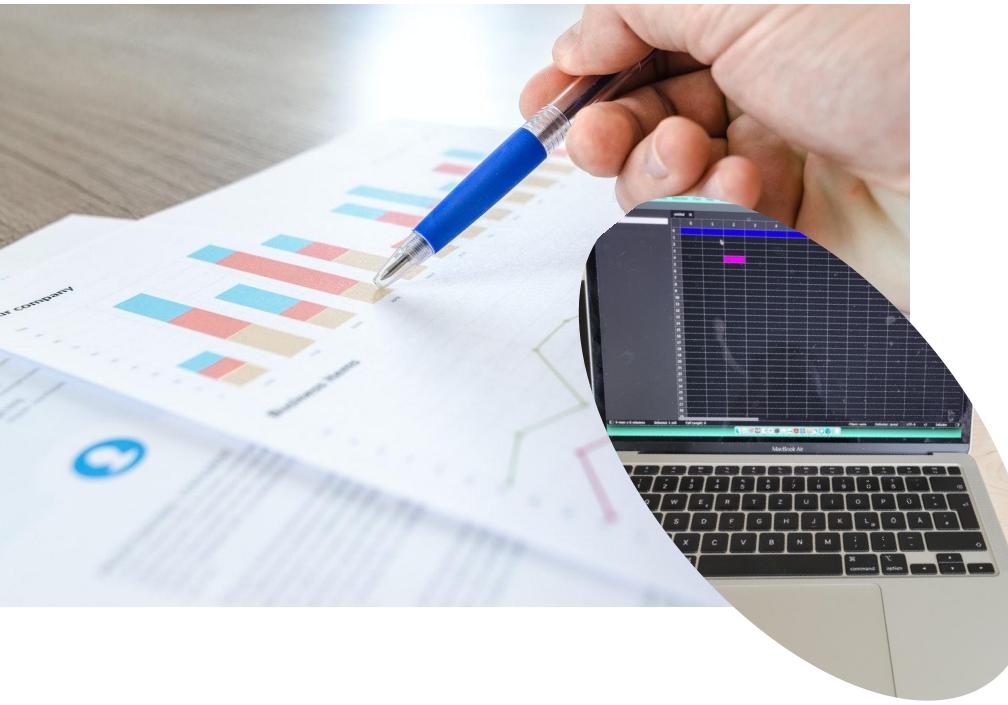
Datenbasiert entscheiden welche Akten langfristig archiviert werden sollen.



Vorgehen bei der Beantwortung von Fragen mit Daten

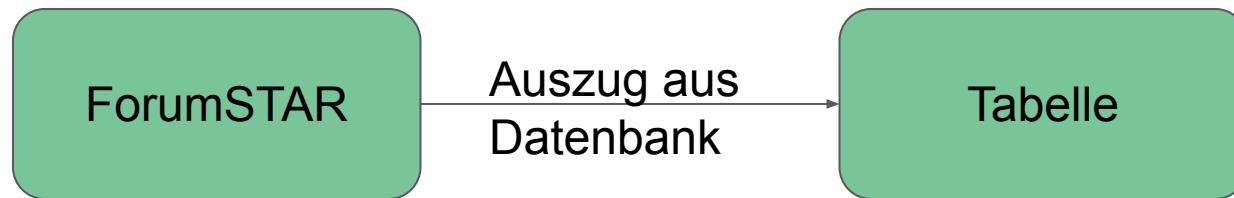


Was sind Daten?



Mit welchen Arten von Daten beschäftigen Sie sich?

Welche Daten gibt es in unserem Projekt?



AZ: 1 F 1/15

Daten betr. Scheidung

Wurde dieses Verfahren nach dem 20.06.2012 eingeleitet? ja nein

Sind beide Ehegatten Iraner? ja nein

örtliche Zuständigkeit [?](#)

F_Sch_ROM III

Heirat: 15.08.1990 **Ehezeitbeginn:** 01.08.1990 [?](#) [§](#) **ZU SchAntrag:** 18.01.2015 **Ehezeitende:** 31.12.2014 [?](#)

vor dem Standesbeamten? Ja Nein

Standesamt: Rosenheim **Staat:** Deutschland [?](#)

HeiratRegNr: 15/1990 **(wenn Heiratsort im Ausland)**

Hatten die Ehegatten zwischen Rechtshängigkeit und letzter mtl. Verhandlung zu irgendeinem Zeitpunkt - gemeinsam ihren gew. Aufenthaltsort in Deutschland? ja nein [?](#)

Haben die Ehegatten eine Rechtswahl getroffen? ja nein

Hat es früher eine Trennungsentscheidung gegeben (Art. 9 ROM III-VO)? ja nein [?](#)

Hatten beide Ehegatten bei Anrufung des Gerichts ihren gew. Aufenthaltsort in Deutschland? ja nein

Staatsangehörigkeit bei ZU des Scheidungsantrags (wegen VA gem. Art. 17 Abs. 3 EGBGB):

<input checked="" type="checkbox"/> Ehemann: Türkei	<input checked="" type="checkbox"/> Ehefrau: Türkei
<input type="checkbox"/> außerdem - nicht "effektiv" -	<input type="checkbox"/> außerdem - nicht "effektiv" -
<input type="checkbox"/> Ehemann ist staatenlos	<input type="checkbox"/> Ehefrau ist staatenlos

War einer der Ehegatten bei ZU des Scheidungsantrags Flüchtling oder Asylberechtigter?

Antragsteller/in <input checked="" type="checkbox"/> Ehefrau <input type="checkbox"/> Ehemann	<input type="checkbox"/> ja <input checked="" type="checkbox"/> nein
VB Antragsgegner/in <input checked="" type="checkbox"/> kein VB oder VB stellt keinen Antrag <input type="checkbox"/> stellt Scheidungsantrag <input type="checkbox"/> beantragt Abweisung Scheidungsantrag	<input type="checkbox"/> widerspricht <input type="checkbox"/> widerspricht nicht <input type="checkbox"/> äußert sich nicht
Antragsgegner/in persönlich <input checked="" type="checkbox"/> stimmt zu <input type="checkbox"/> widerspricht	<input type="checkbox"/> ja <input type="checkbox"/> nein

Soll der Verfahrensteil Scheidung begründet werden? ja nein

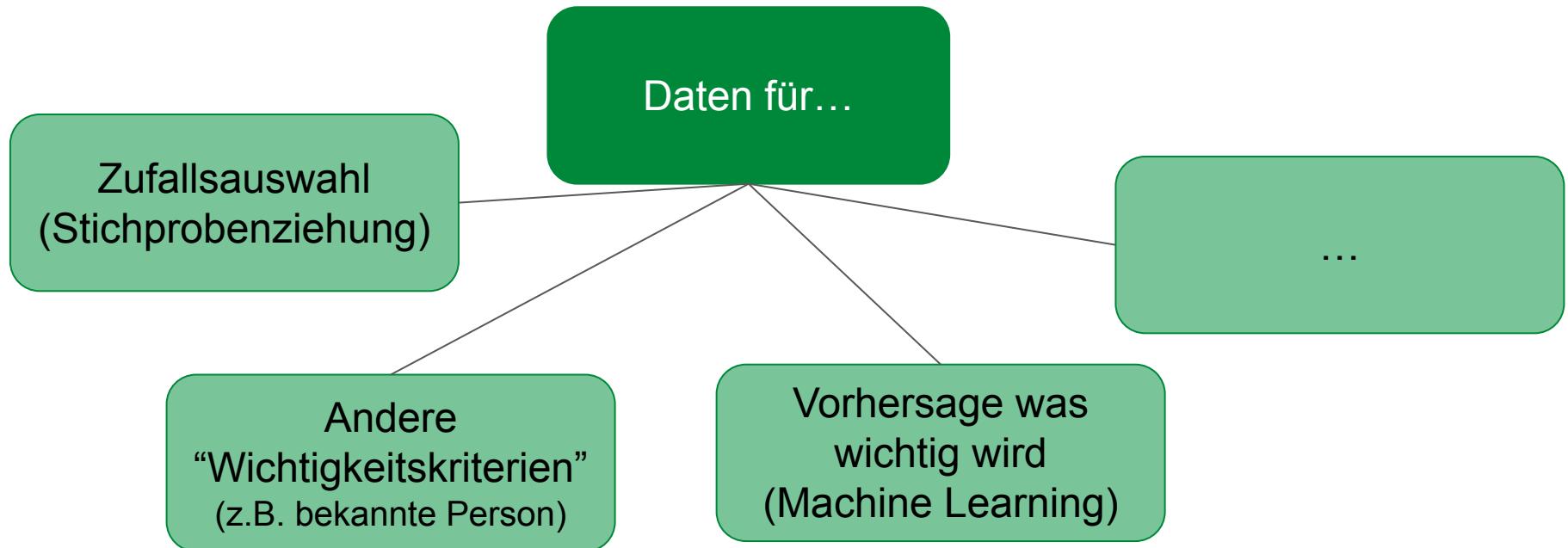
Feststellungen des Gerichts zur Dauer des Getrenntlebens weniger als 1 Jahr (unzumutbare Härte) mindestens 1 Jahr

[Generelle Vorgaben ändern](#)

Weiter [E](#)

DE 15:46 18.01.2015

Welche Daten könnte man noch nutzen?



Daten: das neue Gold

Was macht Daten so wertvoll?
(auch in unserem Projekt)

Zeitersparnis
(Automatisierung)

Verbesserung der
Archivierung

Bürger:innen glücklich
machen

Forschung
verbessern

...

Fragen?

Daten sammeln und verstehen

Selbst erhoben

Von anderen
Organisationen

Aus dem Internet

Woher können Daten kommen?

Von Kollaborations-
Partnern

...

Open Data

<https://bund.dev/apis>

bundDEV
VERWALTUNG DIGITAL



Egal ob Sie als Privatperson eine Open-Source-Anwendung bauen oder als Unternehmen Open-Data in Ihre Produkte integrieren wollen. Unsere Schnittstellen stehen der gesamten Gesellschaft offen.

<https://www.opendata.bayern>

Open Data
Bayern

Offene Daten aus Bayern

Hier finden Datenbegeisterte freie Datensätze und die Verwaltung Unterstützung, um mehr Daten zu teilen. Damit schaffen wir gemeinsam - Verwaltung, Unternehmen, aber auch Wissenschaft und Zivilgesellschaft - Mehrwert für uns alle.

Daten entdecken

[Daten liefern](#)

F A I R

FINDABLE



ACCESSIBLE



INTEROPERABLE



REUSABLE



Where is
the data?

How can
I access it

How can I
read and work
with it?

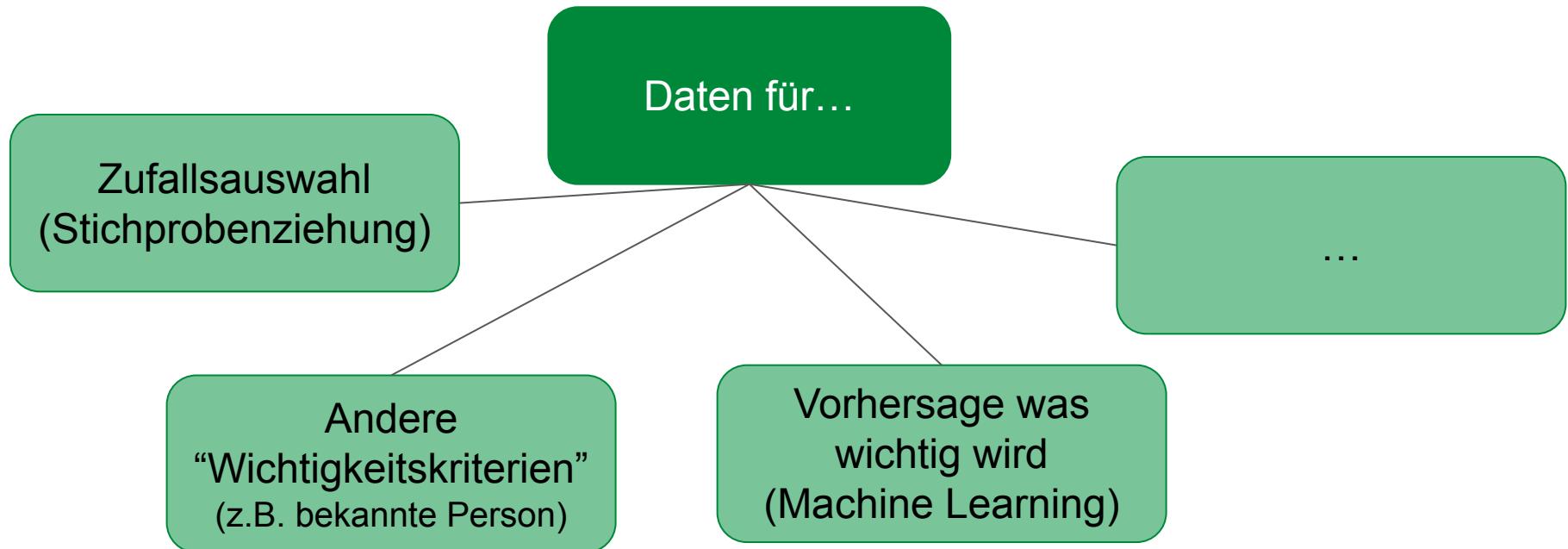
How can I
reuse it?

F A I R

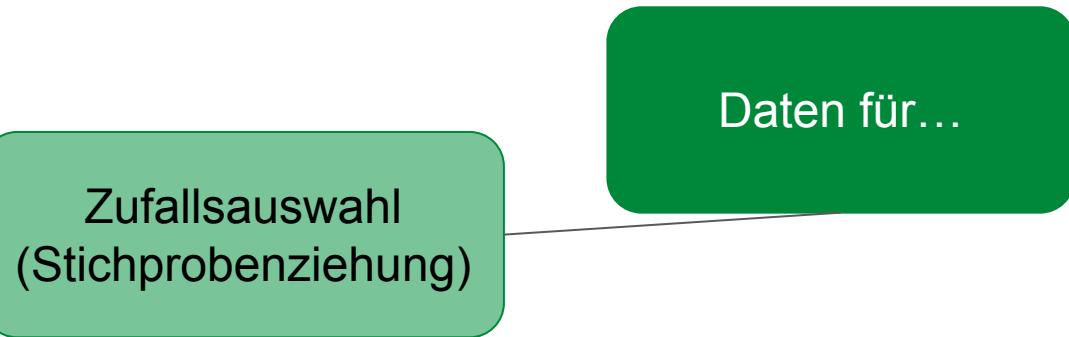
≠

OPEN

Woher könnten Daten kommen?



Woher könnten Daten kommen?



Wie entstehen die verfügbaren Daten?



Was ist eine Datenbank?

- Sammlung von Daten über Entitäten
- Datenmodell von “Relationalen Datenbankmanagementsystemen” (RDBMS):
 - Daten sind in **Tabellen** gespeichert
 - Beziehungen zwischen Tabellen und deren Spalten werden in einem **Schema** definiert
 - RDBMS verwenden die **Structured Query Language** (SQL) um Nutzer mit den Daten interagieren zu lassen
- RDBMS sind weit verbreitet, da
 - viele Nutzer **gleichzeitig** Daten abrufen können
 - Daten bleiben jederzeit **konsistent** und fehlerfrei (*ACID*), selbst bei gleichzeitigen Updates und bei Netzwerkfehlern
- Alternativ: Vielfältige NoSQL-Datenbanken

Ein simples Datenbank-Schema für Gerichtsverfahren

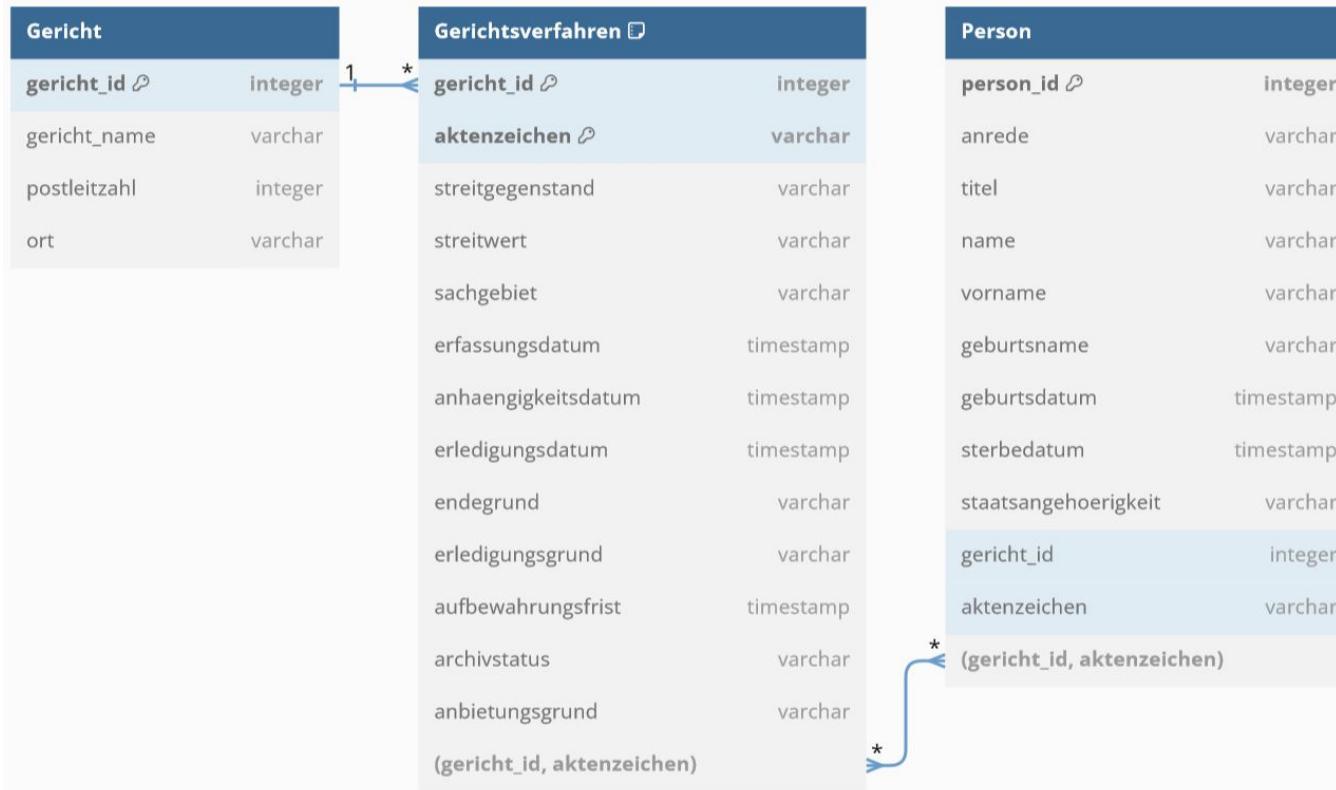


Tabelle: Akten

```
1 SELECT Gericht, Aktenzeichen, Verfahrensstatus, Kurzrubrum, "Streitwert in EURO", Gesamtstreitgegenstand, Streitgege  
2 FROM cases;
```

	Gericht	Aktenzeichen	Verfahrensstatus	Kurzrubrum	Streitwe	Gesamtst
1	Amtsgericht Erlangen	1 C 100/18	weggelegt	ANONYMISIERT	7.920 [NULL]	
2	Amtsgericht Erlangen	1 C 105/18	weggelegt	ANONYMISIERT	1.777,15 [NULL]	
3	Amtsgericht Erlangen	1 C 11/18	weggelegt			
4	Amtsgericht Erlangen	1 C 16/18	weggelegt			
5	Amtsgericht Erlangen	1 C 21/18	weggelegt			
6	Amtsgericht Erlangen	1 C 27/18	weggelegt			
7	Amtsgericht Erlangen	1 C 3/18	weggelegt			
8	Amtsgericht Erlangen	1 C 7/18	weggelegt			

Tabelle: Schema

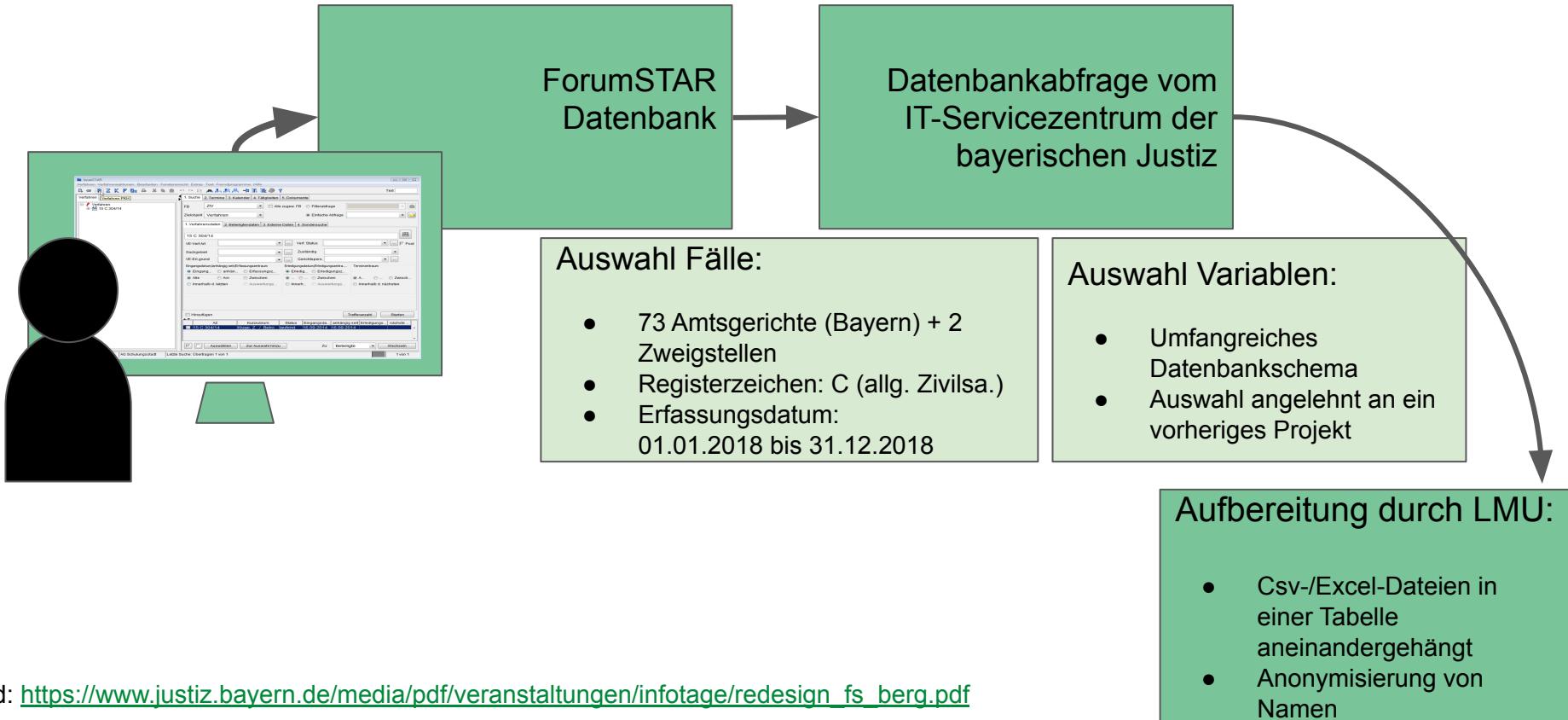
```
5 SELECT * FROM sqlite_schema;
```

	type	name	tbl_name	rootpage	sql
1	table	sqlite_stat1	sqlite_stat1	121	CREATE TABLE sqlite_stat1(tbl,idx,stat)
2	table	sqlite_stat4	sqlite_stat4	122	CREATE TABLE sqlite_stat4(tbl,idx,neq,nlt,ndlt,sample)
3	table	cases	cases	2	CREATE TABLE `cases`(`Gericht` TEXT,`Aktenzeichen` TEXT,`Verfahrensstatus` TEX
4	table	people	people	123	CREATE TABLE `people`(`Gericht` TEXT,`Aktenzeichen` TEXT,`Verfahrensbeteiligu

Tabelle: Personen

	Gericht	Aktenzeichen	Verfahrensbeteiligungsart	Anrede
1	Amtsgericht Erlangen	1 C 100/18	Kläger	Herr
2	Amtsgericht Erlangen	1 C 100/18	Beklagter	Herr
3	Amtsgericht Erlangen	1 C 100/18	Dritter	Frau
4	Amtsgericht Erlangen	1 C 100/18	Dritter	Herr
5	Amtsgericht Erlangen	1 C 100/18	Dritter	Herr
6	Amtsgericht Erlangen	1 C 100/18	Dritter	Frau

Wie entstehen die verfügbaren Daten?



Warum ist es wichtig
zu wissen, wie die
Daten entstanden
sind?

Welche externen Datenquellen können genutzt werden?

Daten für...

Andere
“Wichtigkeitskriterien”
(z.B. bekannte Person)

```
1 SELECT ?person ?personLabel ?birthDate ?description
2 WHERE {
3   ?person wdt:P27 wd:Q183; # German citizenship
4   wdt:P569 ?birthDate. # Birth date property
5   FILTER (YEAR(?birthDate) > 1985). # Filter for birth after 1950
6
7   OPTIONAL {
8     ?person schema:description ?description. # Description property
9     FILTER (LANG(?description) = "de"). # Filter for English or German language description
10  }
11  SERVICE wikibase:label { bd:serviceParam wikibase:language "de". }
12 }
```



Table ▾

26949 Ergebnisse in 23922 ms

</> Code

Herunterladen ▾

Link ▾

Search

person	personLabel	birthDate	description
wd:Q88991	Lena Malkus	6. August 1993	deutsche Weitspringerin
wd:Q89097	Carmen Klaschka	8. Januar 1987	deutsche Tennisspielerin
wd:Q89128	Laura Siegemund	4. März 1988	deutsche Tennisspielerin

Teamarbeit 2

- Was wären optimale Möglichkeiten für die Auswahl (basierend auf Daten)?
 - Was müssten wir dafür digitalisieren?
 - Wie würde eine optimale Welt hier aussehen?
- Warum archivieren wir nicht einfach alle Akten?
- Falls noch Zeit ist: Gibt es bei Ihnen ähnliche Fragen/Projekte, die man mit Hilfe von Daten angeht/angehen könnte?



Überblick Arbeitsumgebung

Kontext

Technischer Rahmen und Zugang

Warum sollten
wir öffentliche
Daten verarbeiten?

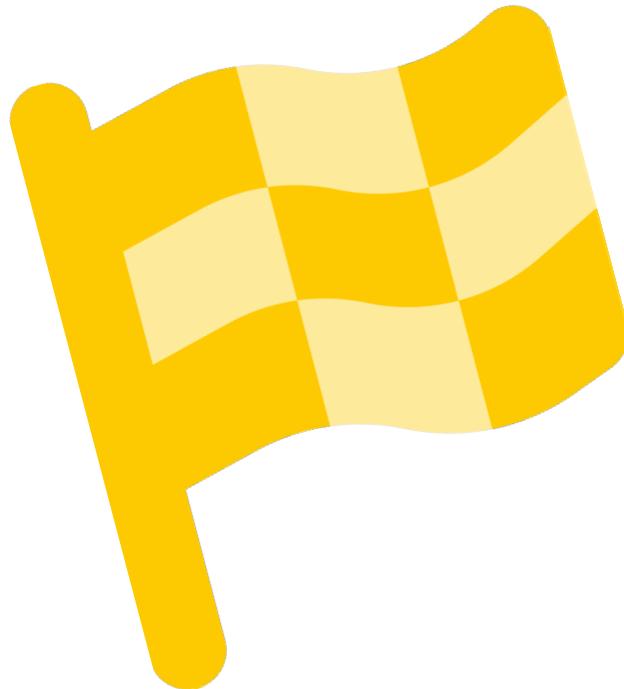
Wie funktioniert
die Analyse
praktisch?



Agenda

In den kommenden **30-40 Minuten**
wollen wir ...

- Ihnen kurz die **Arbeitsumgebung** vorstellen, die wir für den gesamten Workshop nutzen werden
- Gemeinsam Ihren **Zugang** (fertig) einrichten





Coleridge Initiative & ADRF

Die Coleridge Initiative bemüht sich um die Verbreitung von evidenzbasierter Verwaltung

- Vermittelt Zugang zu mehr als 400 vertraulichen, fein aufgelösten Datensätzen von 50+ US-Institutionen
- Bereitet Daten auf und schafft, wo nötig, Querverbindungen (data linkage)
- Entwickelt die technische Plattform für die sichere Arbeit mit den Daten

Rahmenkonzept: Five Safes



Safe
projects



Safe
people



Safe
settings



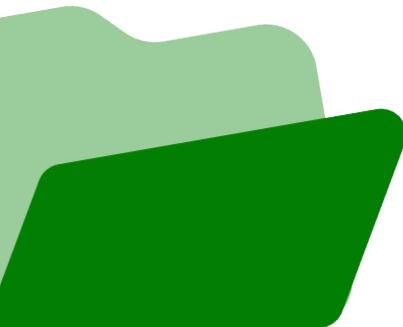
Safe
data



Safe
export

Safe projects

Geprüfte und **zugelassene**
Vorhaben mit nachgewiesenem
öffentlichen Nutzen sowie
Zustimmung der DatengeberIn



Safe people

Individuen mit **Mandat** einer
öffentlichen Institution,
entsprechender **Qualifikation** bzw.
Weiterbildung und Zustimmung zur
'data use agreement'



Safe data

Detaillierte Daten sind häufig
persönlichkeitsrelevant

- Prüfung der Daten auf Validität und
technische Integrität
- Ggf. Deidentifizierung von
Variablen oder Verwendung
synthetischer Daten

01

10

Safe settings

Geschützter Rahmen für die
Datenverarbeitung, der Daten
vor äußerem Zugriff und
unintendierter
Veröffentlichung schützt



Safe settings

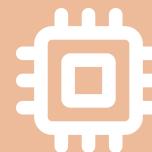
Geschützter Rahmen für die Datenverarbeitung, der Daten vor äußerem Zugriff und unintendierter Veröffentlichung schützt



Zugangs-Kontrolle



Fernsteuerung per KVM



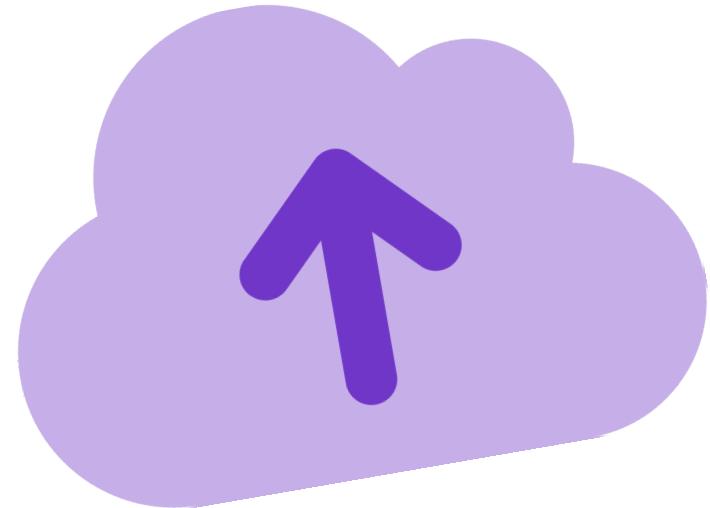
Kontrollierter Datenexport



Safe export

Informationen dürfen die Plattform
nur nach Kontrolle verlassen

- Statistiken dürfen keinen Rückschluss auf individuelle Daten erlauben
- Ein 'export request' wird zunächst von Ihnen selbst, dann von den MitarbeiterInnen am ADRF geprüft



Data Stewardship



Die technischen Schutzmaßnahmen sind nur so wirksam wie das Verhalten der NutzerInnen

- **Teilen Sie Ihren Zugang nicht**
 - Geben Sie Ihre Zugangsdaten nicht weiter
 - Nutzen Sie das ADRF nur aus einer sicheren Umgebung
 - Lassen Sie Ihren Computer nicht unbeaufsichtigt
- **Exfiltrieren Sie keine Daten aus der ADRF**
 - Machen Sie keine schriftlichen Notizen
 - Nehmen Sie keine Screenshots
 - Übertragen Sie Ihren Bildschirminhalt nicht per Zoom

Verletzungen der 'five safes'



Coleridge nimmt Verletzungen der 'five safes' sehr ernst, und behalten sich Reaktionen vor

- Sofortiger **Verlust des Zugangs** zu Daten
- **Ausschluss** von zukünftiger Nutzung
- **Juristische Konsequenzen** lt. 'Data Use Agreement'



Verhalten und Nutzung werden im ADRF protokolliert und können überwacht werden.

Angemessene Nutzung

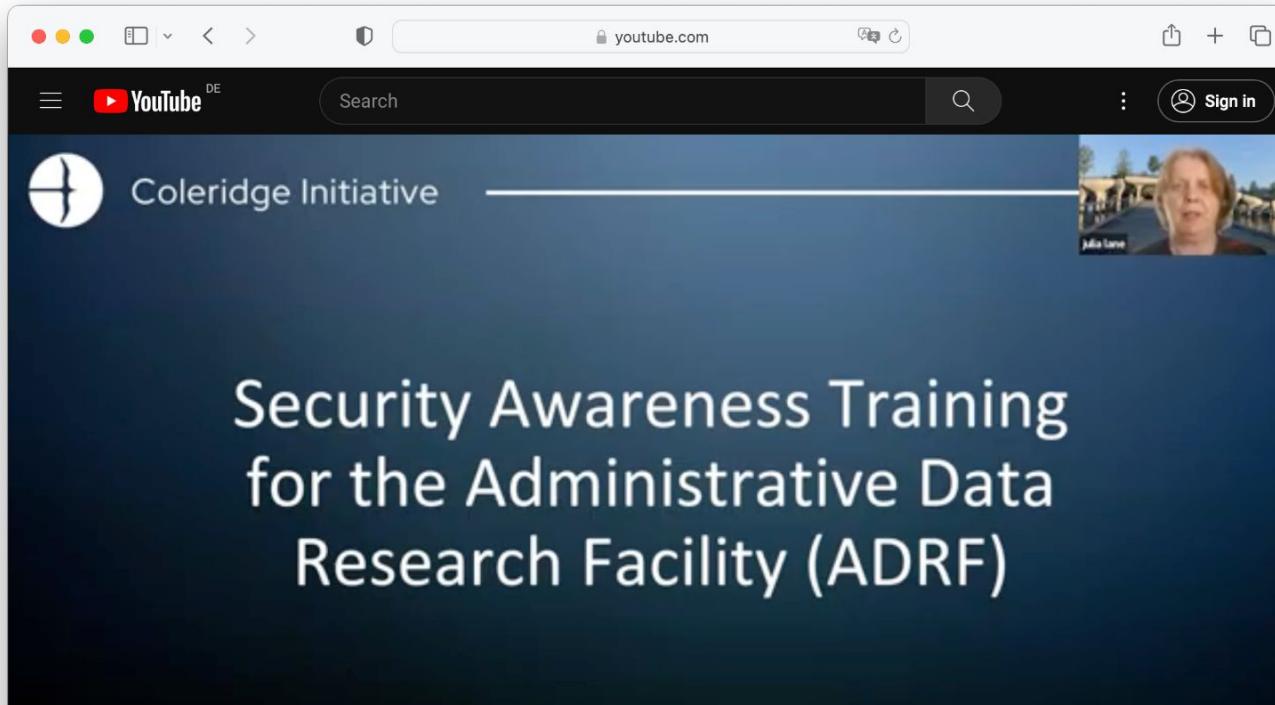
Bitte betrachten Sie den Zugang zu Daten als Privileg, und behandeln Sie sie als wären es Ihre eigenen

- Erstellen Sie Ihre **Notizen und schriftliche Ergebnisse innerhalb der ADRF**
- **Teilen Sie Informationen** bis zu deren Export **nur innerhalb des Projekts** mit zugelassenen NutzerInnen
- Behandeln Sie Ihre **Login-Daten** vertraulich
- Wir besprechen den **Export von Ergebnissen**



Weitere Informationen

<https://www.youtube.com/watch?v=GFXjxBblzic>



Demo



Was bisher geschah ...

Sie sollten eine Email von Coleridge erhalten haben, die einen Nutzernamen sowie den Zugangslink zu Okta enthält

The screenshot shows an email inbox with a single message selected. The message is from 'Okta <noreply@okta.com>' to 'felix.henninger@stat.uni-muenchen.de' with the subject 'Coleridge Initiative - Welcome to the ADRF!'. The email body contains the following text:

Welcome to the Administrative Data Research Facility

Hi Felix,

The Coleridge Initiative is using the FedRAMP Authorized Okta Identity Provider to manage your access to the ADRF. This means you can use a single username/password with multi-factor authentication to sign in and gain access to your ADRF Projects and Applications. Watch this short video to learn more: <https://www.okta.com/intro-to-okta/>

Access to the ADRF and all tools available within your ADRF workspace is provided and managed by the Coleridge Initiative, regardless of your agency affiliation.

The ADRF system administrator has created an Okta user account for you. Your ADRF Okta username is **Felix.Henning@adrft.net**

Click the following link to activate your account:

[Activate ADRF Okta Account](#)

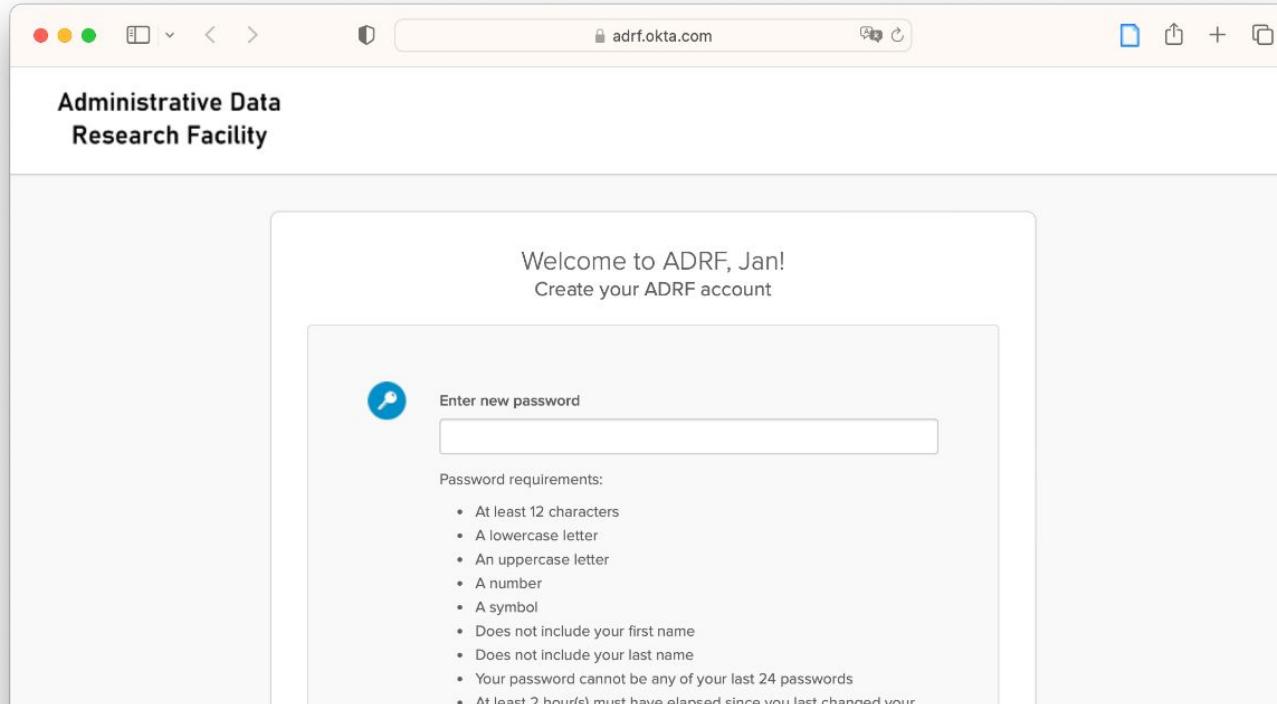
This link expires in 7 days.

You will be prompted to select a strong password and set up multi-factor authentication with at least one of the following factors:

- *Okta Verify Mobile App* - allows for "push" notifications to your mobile device
- *SMS Authentication* - a text message with a numeric code is sent to your mobile device
- *Voice Call Authentication* - a call is placed to your phone number and a confirmation code is provided over the call

Registrierung

Auf Okta können Sie einen Zugang anlegen



Optionen zur Passwort-wiederherstellung

Repeat new password



Choose a forgot password question

What is the food you least liked as a child?



Answer



Add a phone number for resetting your password or unlocking your account using SMS (optional)

Okta can send you a text message with a recovery code. This feature is useful when you don't have access to your email.

Add Phone Number



Add a phone number for resetting your password or unlocking your account using Voice Call (optional)

Okta can call you and provide a recovery code. This feature is useful when you don't have access to your email.

Add Phone Number



Click a picture to choose a security image

Your security image gives you additional assurance that you are logging into Okta, and not a fraudulent website.

Sicherheits-
Bild
(wird beim
login
angezeigt)

Okta can call you and provide a recovery code. This feature is useful when you don't have access to your email.

[+ Add Phone Number](#)



Click a picture to choose a security image

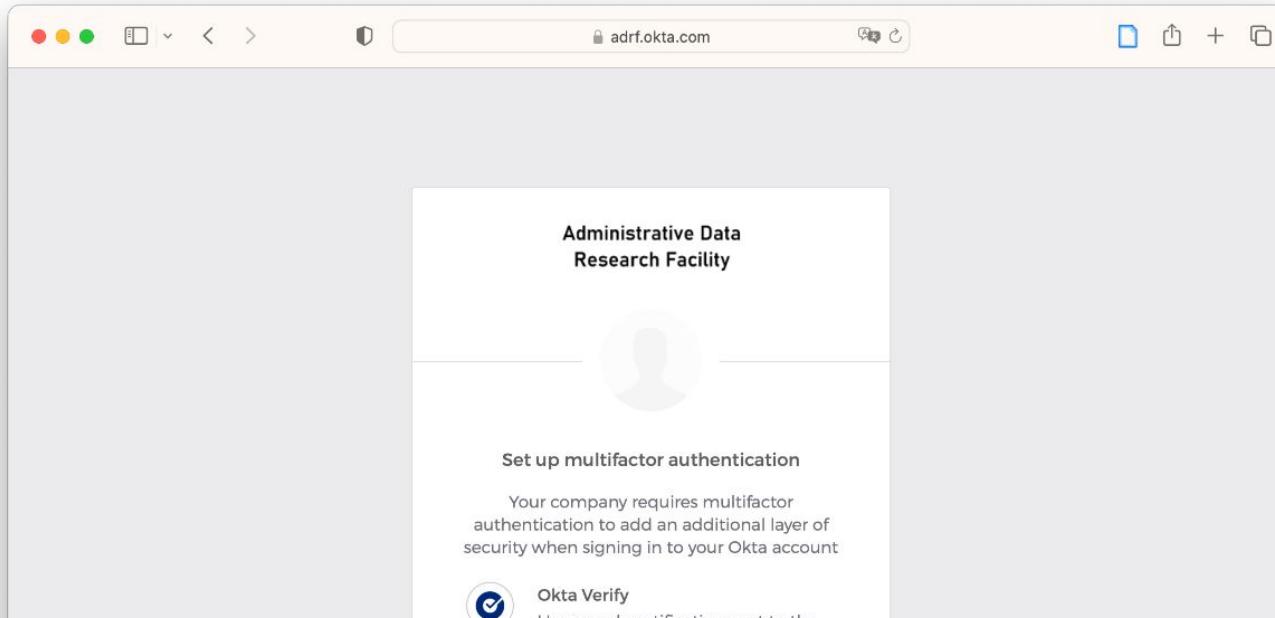
Your security Image gives you additional assurance that you are logging into Okta, and not a fraudulent website.



[Create My Account](#)

Zwei-Faktor-Authentifizierung

Jeder Login wird über einen unabhängigen Kanal bestätigt



Optionen für
den zweiten
Faktor
(App, SMS,
Anruf, Email)

Administrative Data Research Facility



Set up multifactor authentication

Your company requires multifactor authentication to add an additional layer of security when signing in to your Okta account



Okta Verify

Use a push notification sent to the mobile app.

[Setup](#)



SMS Authentication

Enter a single-use code sent to your mobile phone.

[Setup](#)



Voice Call Authentication

Use a phone to authenticate by following voice instructions.

[Setup](#)



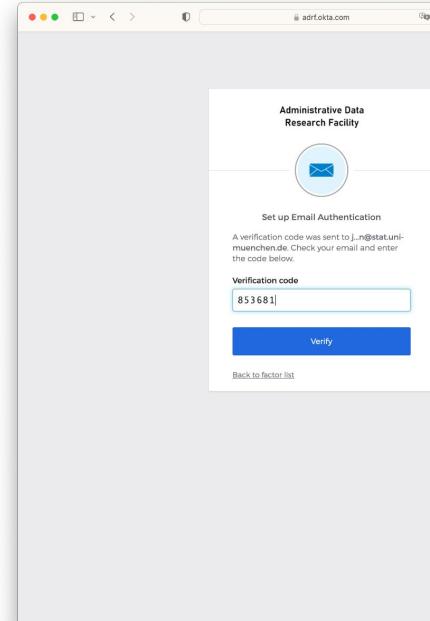
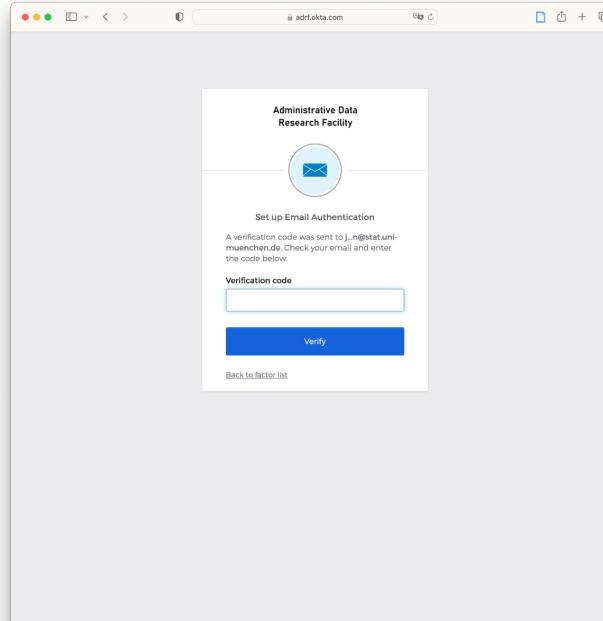
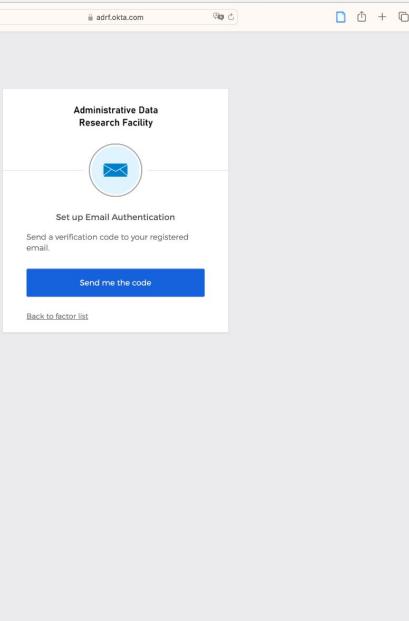
Email Authentication

Enter a verification code sent to your email.

[Setup](#)

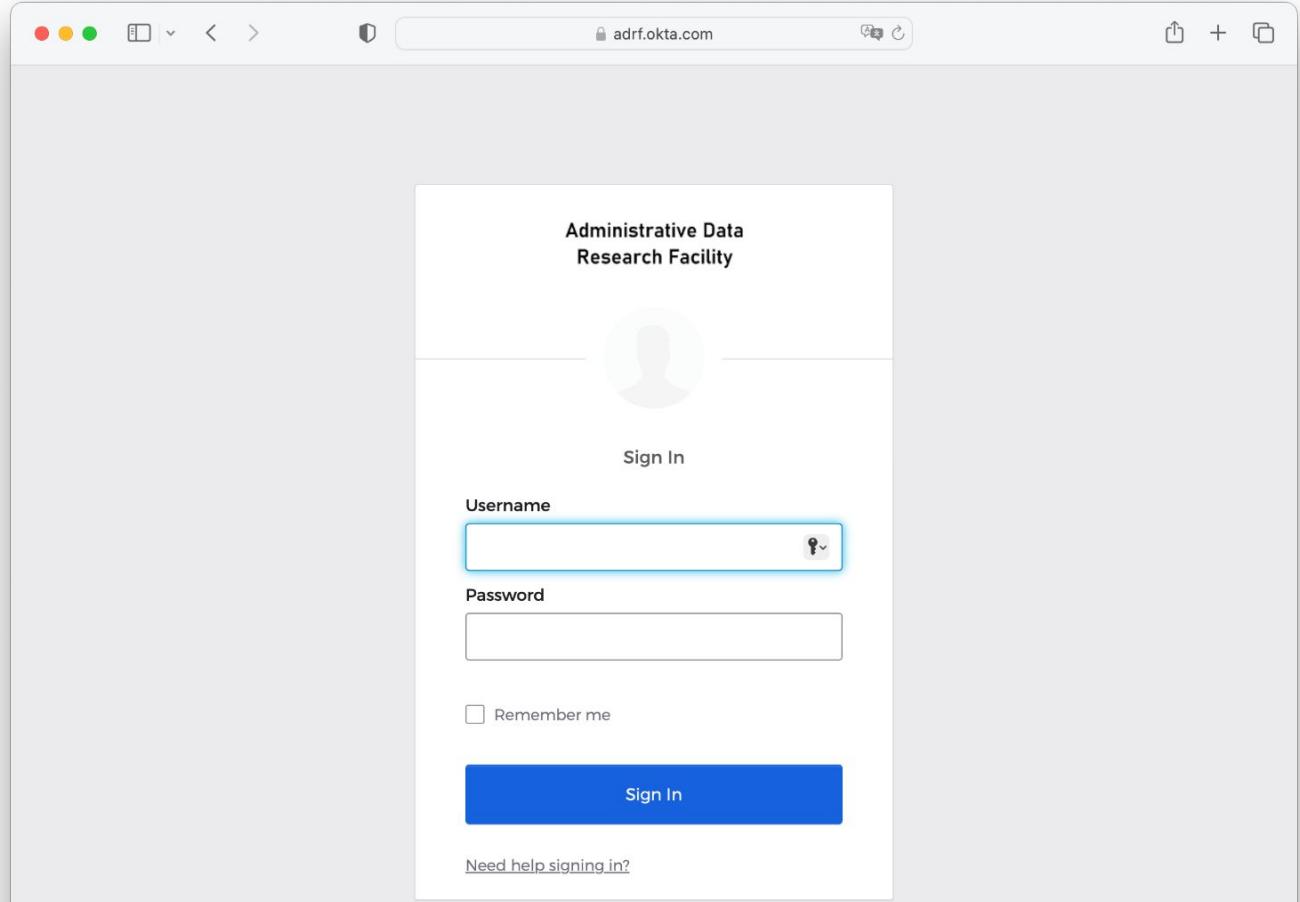
Login

Über Ihren gewählten zweiten Faktor erhalten
Sie einen Freischaltungscode



Login

Zugang erhalten Sie
immer über die
Adresse
<https://adrf.okta.com>



The screenshot shows a web browser window with the URL adrf.okta.com in the address bar. The page title is "Administrative Data Research Facility". The login form includes fields for "Username" and "Password", a "Remember me" checkbox, and a "Sign In" button. A link for "Need help signing in?" is at the bottom.

adrf.okta.com

Administrative Data
Research Facility

Sign In

Username

Password

Remember me

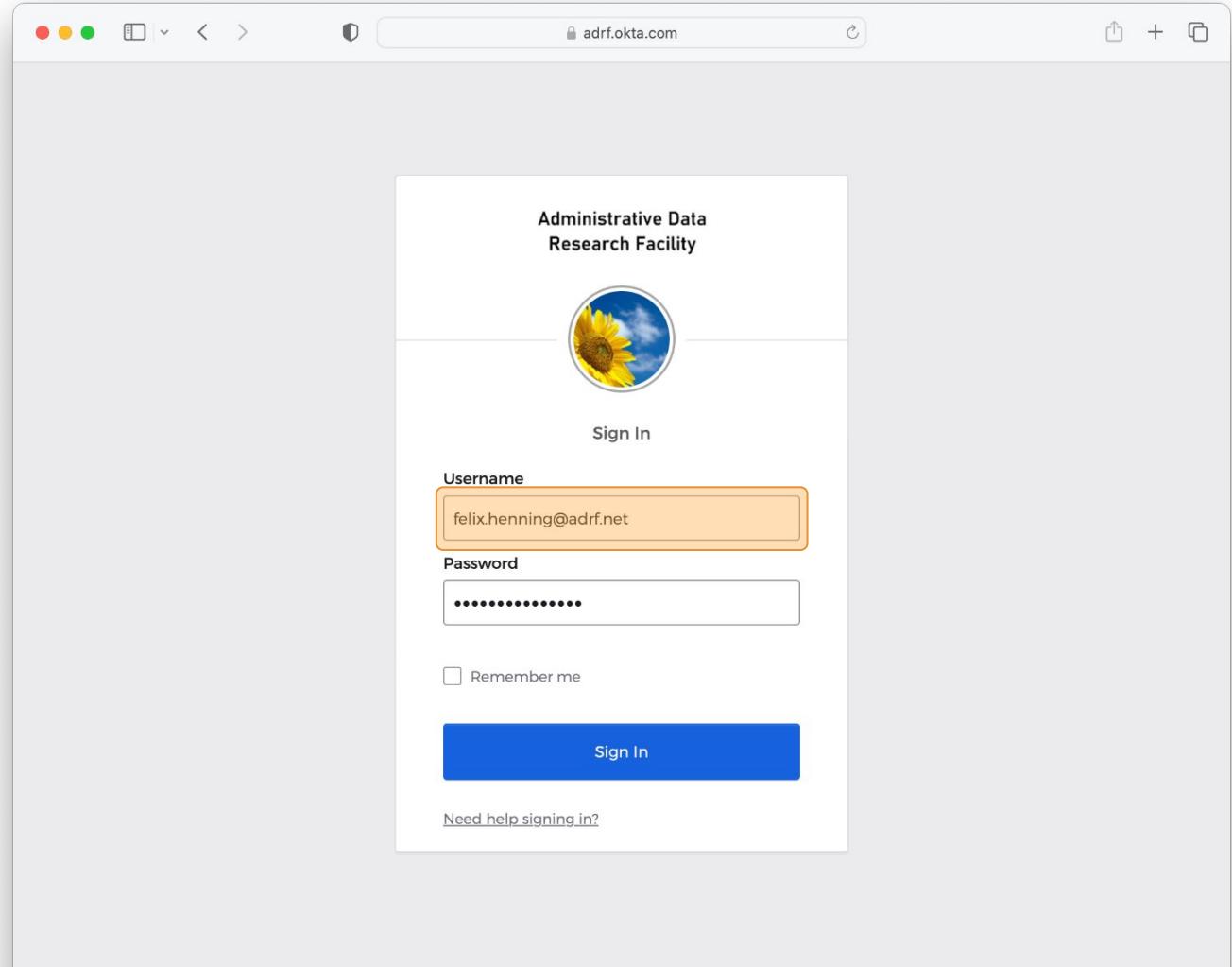
Sign In

[Need help signing in?](#)

Login

Zugang erhalten Sie
immer über die
Adresse
<https://adrf.okta.com>

Bitte loggen Sie sich
mit der **vom ADRF**
vergebenen
Email-Adresse ein



The screenshot shows a web browser window with the URL adrft.okta.com in the address bar. The page itself is titled "Administrative Data Research Facility" and features a circular logo with a sunflower against a blue sky with clouds. Below the logo is a "Sign In" button. The "Username" field is populated with "felix.henning@adrft.net" and has an orange border, indicating it is selected or highlighted. The "Password" field contains several dots. There is a "Remember me" checkbox followed by a "Sign In" button in blue. At the bottom, there is a link "Need help signing in?".

Administrative Data
Research Facility

Sign In

Username

felix.henning@adrft.net

Password

.....

Remember me

Sign In

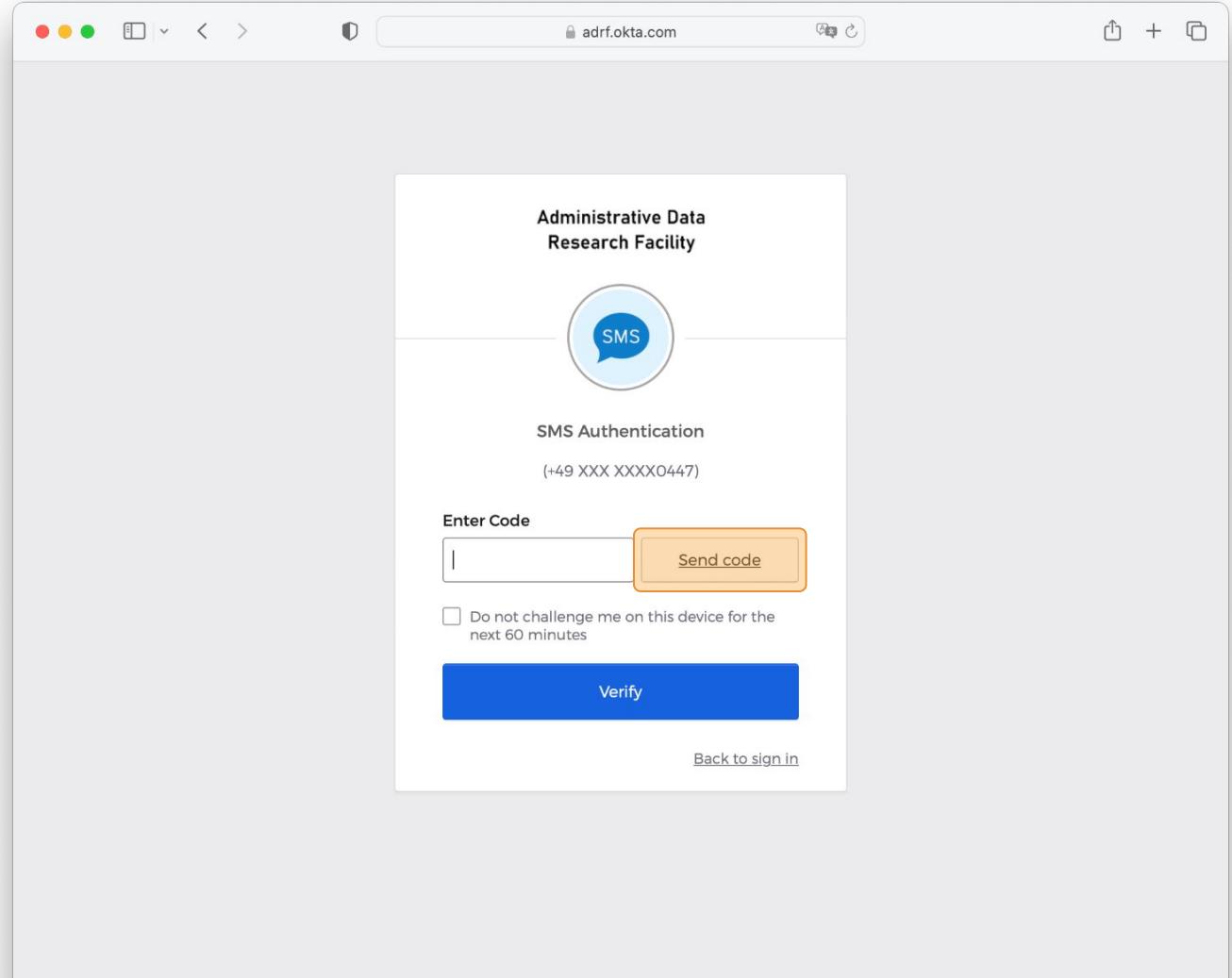
[Need help signing in?](#)

Login

Zugang erhalten Sie
immer über die
Adresse
<https://adrf.okta.com>

Bitte loggen Sie sich
mit der **vom ADRF**
vergebenen
Email-Adresse ein.

Bei jedem login wird
Ihr **zweiter Faktor**
geprüft.

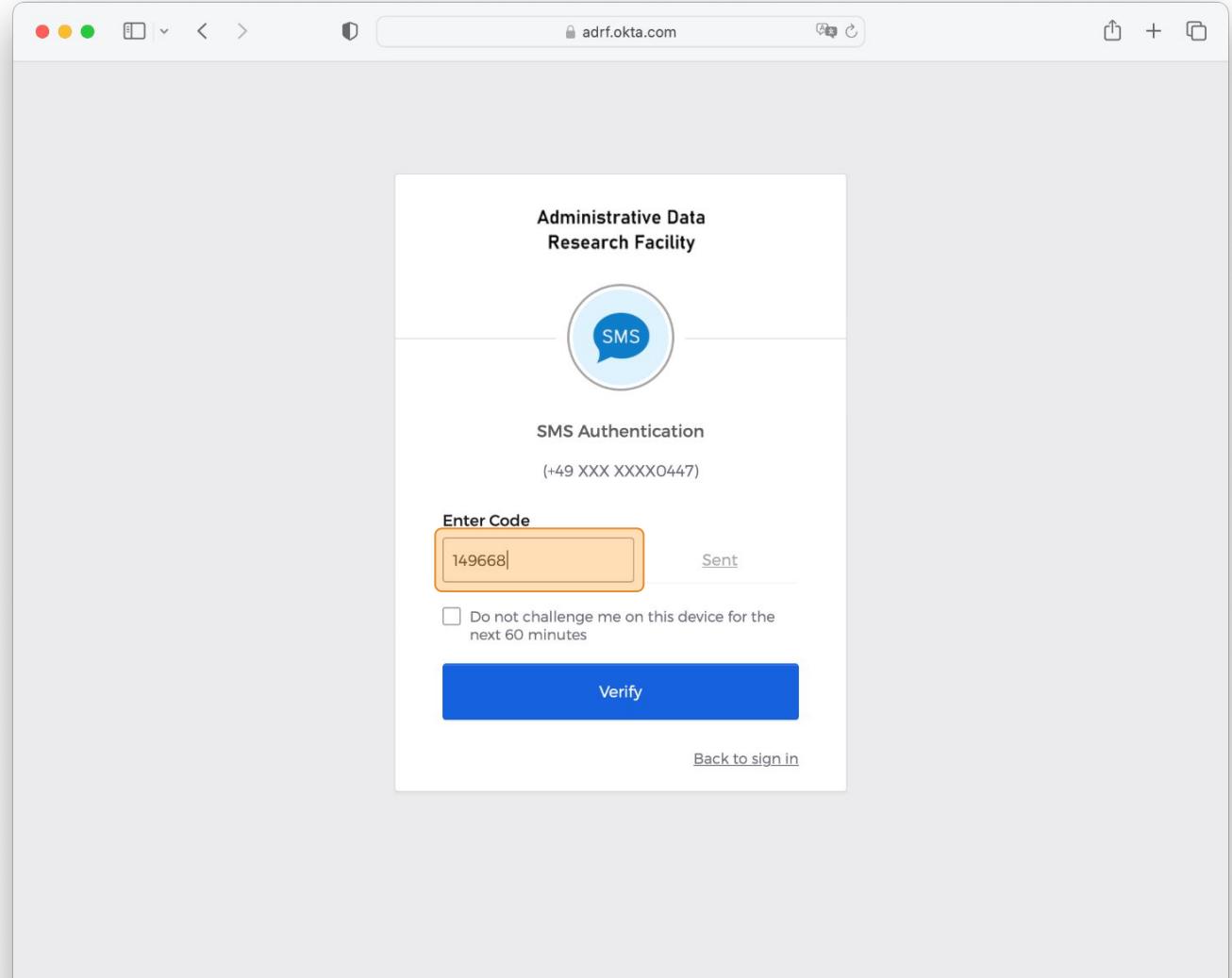


Login

Zugang erhalten Sie
immer über die
Adresse
<https://adrf.okta.com>

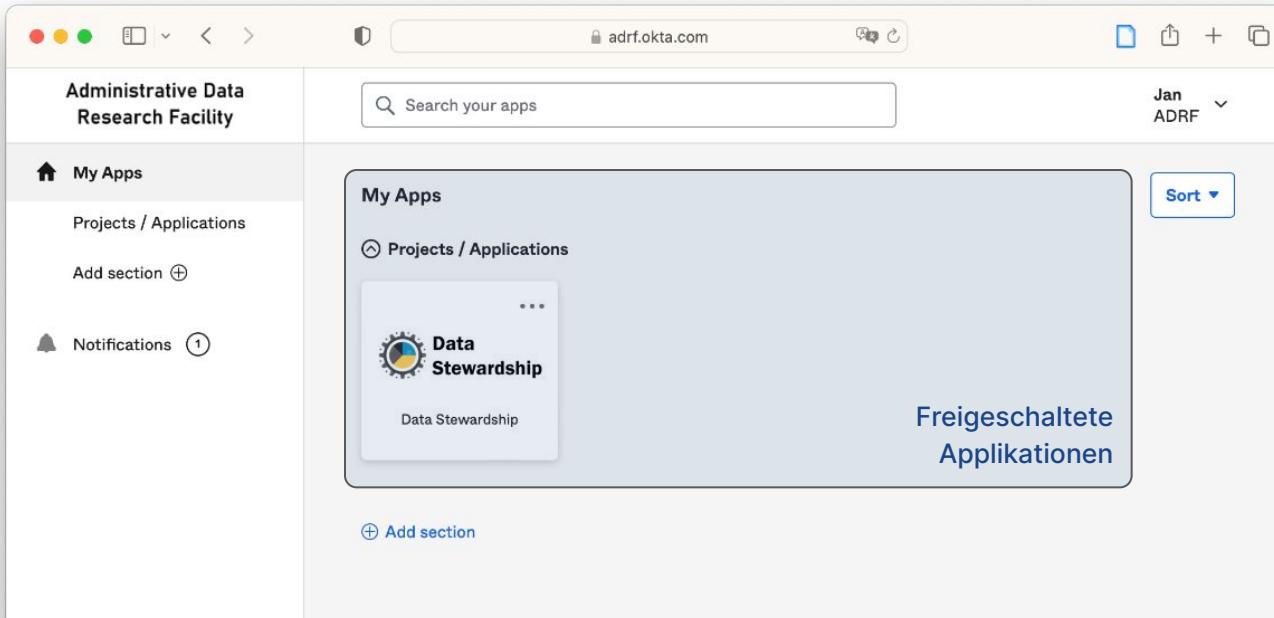
Bitte loggen Sie sich
mit der **vom ADRF**
vergebenen
Email-Adresse ein.

Bei jedem login wird
Ihr **zweiter Faktor**
geprüft.



Applikationen

Nach dem Login stehen Ihnen die Applikationen des ADRF zur Verfügung



The screenshot shows a web browser window for adrf.okta.com. The title bar includes standard OS X window controls (red, yellow, green, close, minimize, maximize) and a search bar. The top right corner shows the user's name "Jan ADRF" and a dropdown arrow. The main content area has a light gray header with the text "Administrative Data Research Facility". On the left, a sidebar menu includes "My Apps" (selected), "Projects / Applications", and "Add section". Below this is a "Notifications" section with a bell icon and a red circle containing the number "1". The main panel displays a "My Apps" section with a "Projects / Applications" card. This card features a circular icon with a gear and a blue/purple gradient, the text "Data Stewardship", and "Data Stewardship" below it. To the right of the card, the text "Freigeschaltete Applikationen" is overlaid in blue. At the bottom of the main panel, there is a "Sort" button with a downward arrow and an "Add section" button with a plus sign.

Data Stewardship-Modul

Prüft die Voraussetzungen für den Zugang zur Datenplattform

The screenshot shows a web browser window titled "My Apps Dashboard | ADRF" with the URL "dsa.adrf.net". The user is logged in as "Jan Simson" and has the option to "Logout". On the left, there are two main navigation items: "Usage Statistics" and "Tasks". The "Tasks" item is selected, indicated by a purple icon and the number "3". To the right of the tasks, it shows "Pending: 3" and "Completed: 0". Below this, a "Filter Here" input field is present. A large callout box highlights the "Tasks" section, listing three items:

- ADRF - Terms of Use**
Assigned: June 8, 2021
- Security Training Quiz**
Assigned: June 17, 2021
- Security Training Video**
Assigned: June 17, 2021

Each task item has a red exclamation mark icon to its right. To the right of the highlighted tasks, the text "Ausstehende Aufgaben" is displayed in blue.

Zugangsvoraussetzungen

Unterzeichnung der Nutzungsbedingungen

Please review the documents below.

DocuSign Envelope ID: B91A8E24-69EE-42A2-B463-03030DF21220

My Apps Dashboard | ADRF Review and sign document(s) | DocuSign FINISH MORE OPTIONS

START

Administrative Data Research Facility (ADRF)
Terms of Use

The agreement below must be signed before you are given access to the Administrative Data Research Facility ("ADRF"). Please email support@coleridgeinitiative.org or ask an ADRF staff member for clarification if you have questions about any of the terms below.

I understand that I am being granted access to confidential information. I understand that my access to the ADRF is limited to the scope of my authorized access. The rules of my access to data for my project are set by the data owner and will be provided to me as part of that access.

I understand that my access to systems and networks owned by the Coleridge Initiative is governed by, and subject to, all relevant law. This may include, but is not limited to, the Privacy Act, 5 U.S.C. § 552a.

Rules of Behavior – Do's and Don'ts for Discussing Data inside the ADRF

It is important to protect the confidential data that is inside the ADRF as communicating with your team-mates. The general rule is that you should never take any exact number out of the ADRF. This means you should never write down or share any number by text, screenshot, or share an image even with a team-mate. The rules have become more complicated now that everything is online, because even though your team-mates are "safe people", and Zoom conversations are password protected and encrypted, we'd rather err on the side of caution when sharing information over Zoom.

This cheat sheet summarizes some of the rules that apply to discussing data before it has been exported from the ADRF and passed the ADRF team's disclosure review. If you are unsure about a specific situation, please ask an instructor.

Exact Numbers

Do not describe a statistic in exact numbers. If you would like to communicate these values while not in person, you can have a private discussion via the projects drive.

Example: If an average within a specific group was 5,000, you would need to convey this average on the projects drive.

Comparing Values

When comparing values, you are permitted to say that one value is more than, less than, or about the same as another. However, you cannot refer to the exact difference between the two numbers.

In practice, you can use pluses and minuses to convey differences between values for data that has not been exported from the ADRF.

Example: "The mean for Group A was roughly the same as the mean for Group B, but these values were both greater than that of Group C."

Percentages/Proportions

Percentages and proportions also cannot be directly mentioned. Instead, you can refer to the percentage/proportion within 25%.

Example: If a proportion was 30%, you could say "The proportion is about 25%" or "The proportion is between 25% and 50%."

1 of 2

Sicherheitsunterweisung

My Apps Dashboard | ADRF Data Stewardship UI Jan Simson | Logout

|Data Stewardship

Usage Statistics Tasks Filter Here Pending: 3 Completed: 0

3 Tasks

Filter Here Pending: 3 Completed: 0

ADRFE - Terms of Use Assigned: June 8, 2021 Security Training Quiz

Security Quiz Video

ADRFE Security Training

Security Awareness Training for the Administrative Data Research Facility (ADRF)

Watch on YouTube https://coleridgeinitiative.org/adrf/

Mark Complete Close

Quiz zur Sicherheit

My Apps Dashboard | ADRF Data Stewardship UI Jan Simson | Logout

|Data Stewardship

Usage Statistics Tasks Filter Here Pending: 3 Completed: 0

3 Tasks

Filter Here Pending: 3 Completed: 0

ADRFE - Terms of Use Assigned: June 8, 2021 Security Training Quiz

Security Quiz Video

Security Training Quiz

Please complete the quiz below. You must answer 5 of 6 questions correctly to pass.

1. It's OK to share data from the ADRF with my colleagues if I swear them to secrecy.

2. It's OK to take a screenshot of a graph from the ADRF and put it in a PowerPoint.

3. It's OK to write down summary information from a table in the ADRF and use it for a presentation.

4. If someone else on the same project doesn't have their login information, it's OK to share my login information with them.

5. Violations of the ADRF and data security protocols may result in:

- a. Termination of access to data
- b. Prevention of future access to data
- c. Penalties outlined in the underlying data use agreement
- d. All of the above

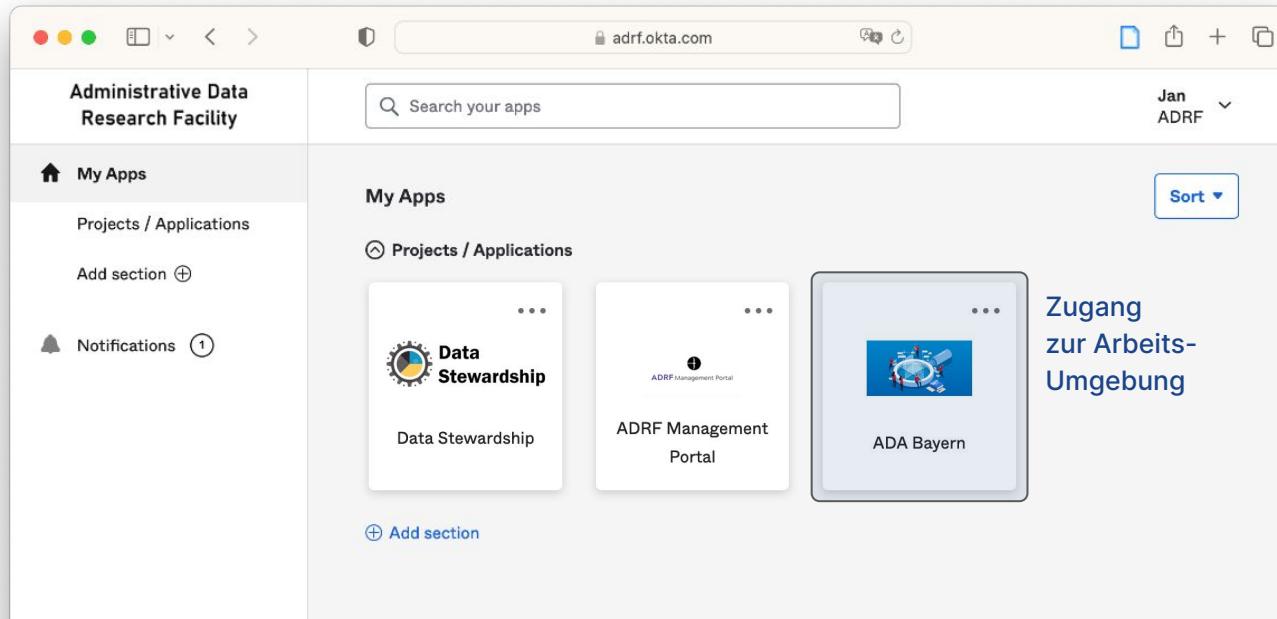
6. If you want to remove output from the ADRF, you have to:

- a. Review the output you want to remove from the ADRF, paying close attention to potential disclosure risks.
- b. Place the output into the export request module and submit your request as per instructions given in the export review module documentation.
- c. Both a and b
- d. Neither a nor b – I can just write the results down and use them without disclosure review

Please review your answers before submitting. You may attempt the quiz as many times as needed to pass, but your answers will be cleared between attempts.

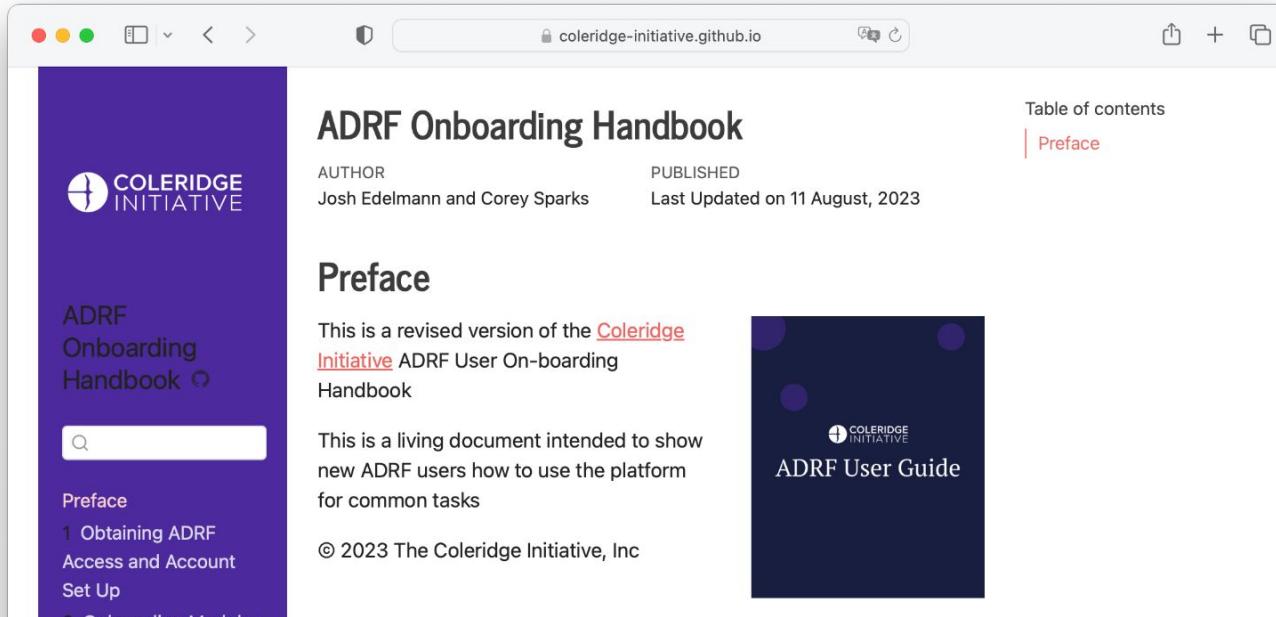
Zugang zur Datenplattform

Nach dem Abschluss der drei Aufgaben erhalten Sie (über Nacht) Zugang zum ADRF



Weitere Informationen ...

.. finden Sie im 'onboarding handbook' unter
https://coleridge-initiative.github.io/adrf_onboarding_handbook/



The screenshot shows a web browser window displaying the 'ADRF Onboarding Handbook' page. The page has a dark purple header with the Coleridge Initiative logo and the title 'ADRF Onboarding Handbook'. Below the header, there's a sidebar with navigation links like 'Preface', '1 Obtaining ADRF Access and Account Set Up', and '2 Coleridge User Guide'. The main content area features the title 'ADRF Onboarding Handbook' and author information ('AUTHOR: Josh Edelmann and Corey Sparks'). It also includes a 'PUBLISHED' section with the date 'Last Updated on 11 August, 2023'. A 'Table of contents' section is visible on the right, with a link to the 'Preface'. The 'Preface' section is currently active, showing its content: 'This is a revised version of the [Coleridge Initiative](#) ADRF User On-boarding Handbook'. Below this, there's a brief description of the document's purpose: 'This is a living document intended to show new ADRF users how to use the platform for common tasks'. At the bottom, there's a copyright notice: '© 2023 The Coleridge Initiative, Inc'. To the right of the main content, there's a small thumbnail image of the 'ADRF User Guide'.

Raum für
Fragen



Ihre Aufgabe

Bitte nehmen Sie sich 15-20 Minuten Zeit, um Ihren Zugang zur ADRF-Plattform freizuschalten

- Legen Sie sich einen Account bei Okta an
- Beantworten Sie im Data Stewardship-Portal die Quizfragen und stimmen Sie der Nutzungsvereinbarung zu
- Die MitarbeiterInnen der Coleridge-Platform schalten Sie danach (bis morgen) frei

