

# Modul 1: Einführung, Problemerkennung und Daten verstehen

Angewandte Datenanalyse für die öffentliche Verwaltung in Bayern (ADA Bayern)  
[www.ada-oeffentliche-verwaltung.de](http://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!

## DOMAIN EXPERT

User, analyst, or leaders with deep subject matter expertise related to the data, its appropriate use, and its limitations

## SYS ADMIN

Team member responsible for defining and maintaining a computation infrastructure that enables large scale computation



## RESEARCHER

Team member with experience applying formal research methods, including survey methodology and statistics

## COMPUTER SCIENTIST

Technically skilled team member with education in computer programming and data processing technology

# 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 Stimmelmayer, 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



- *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

# 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



Smarte Gesellschaftspolitik bedeutet, Daten innovativ, verantwortungsvoll und gemeinwohlorientiert zu nutzen

© BMFSFJ



# 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?

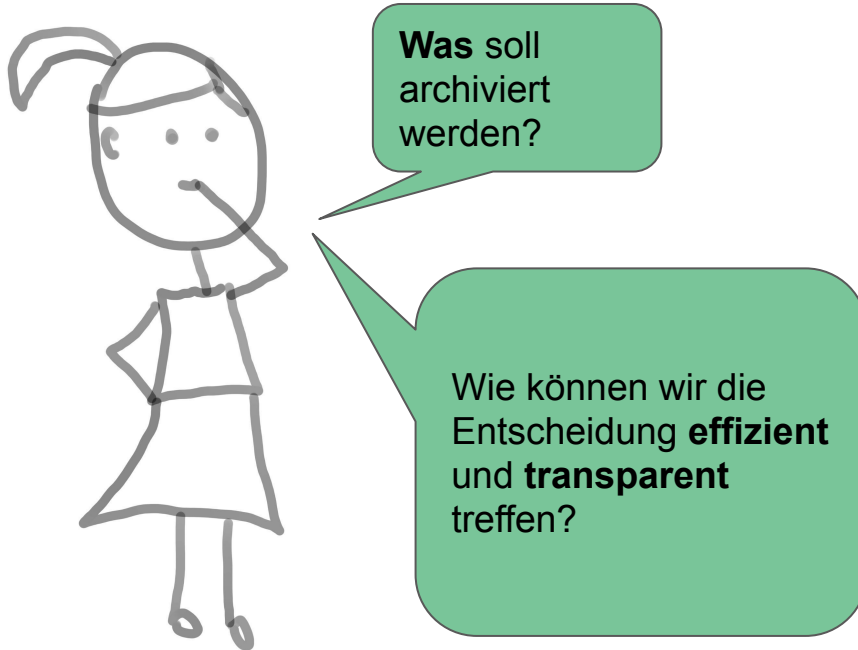
Es gibt Berichte von Sozialarbeitern. Wo müssen wir intervenieren?

Sollen wir diese Maßnahme nochmal machen?

Was soll archiviert werden?

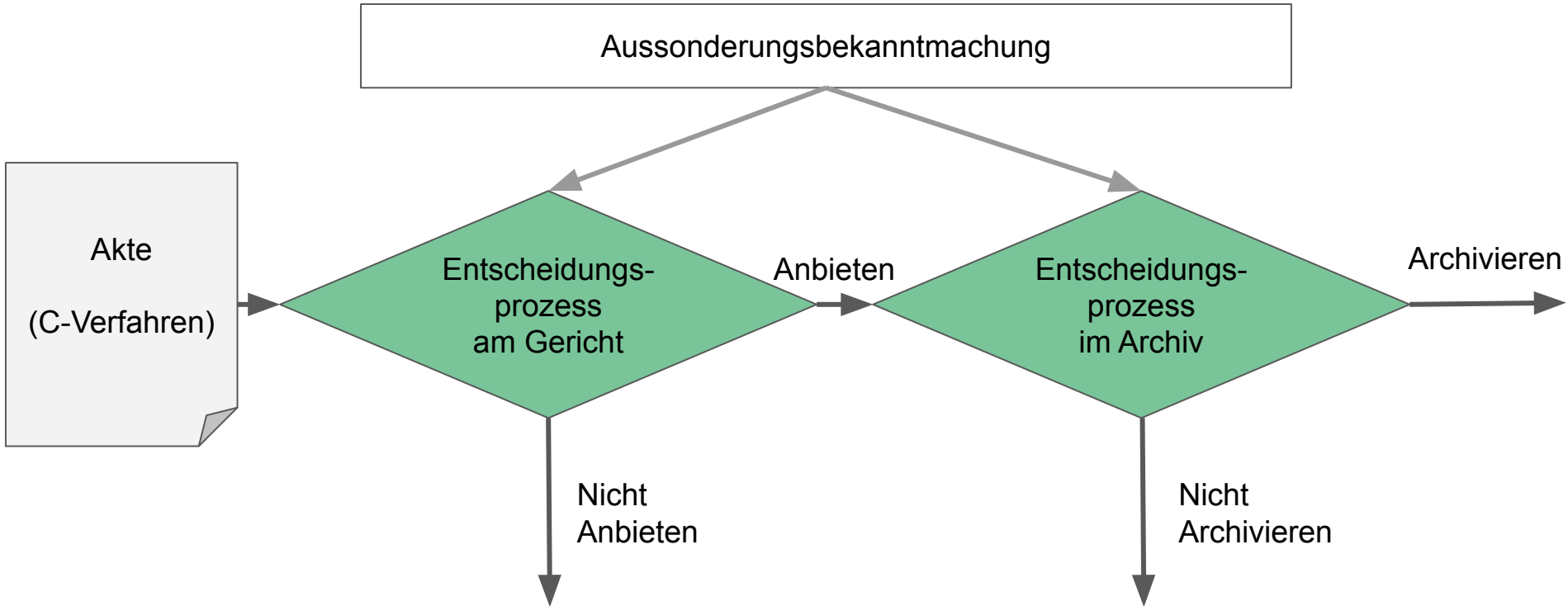
# Was ist eigentlich die Frage? Problemstellung definieren.

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>

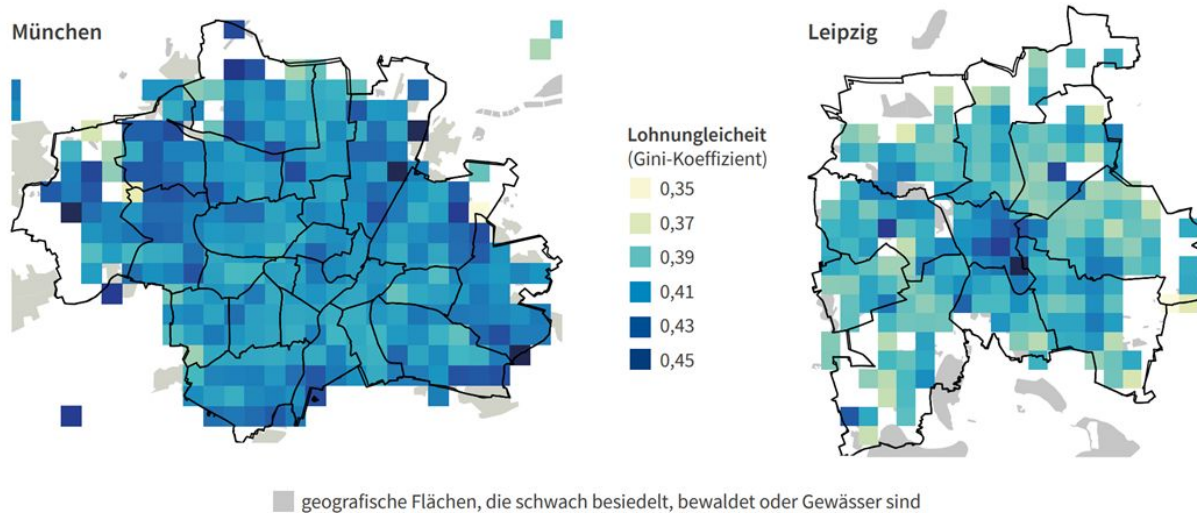
- Kooperation: Johnson County, Kansas  
Carnegie Mellon University
- Ziel: Spirale unbehandelter psychiatrischer Erkrankungen und Verhaftungen zu durchbrechen.
- KI-gestützte Vorhersage der Wahrscheinlichkeit ins Gefängnis zu kommen
- Personen mit hohem Risikowert werden je nach vorhandenen Ressourcen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



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

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.

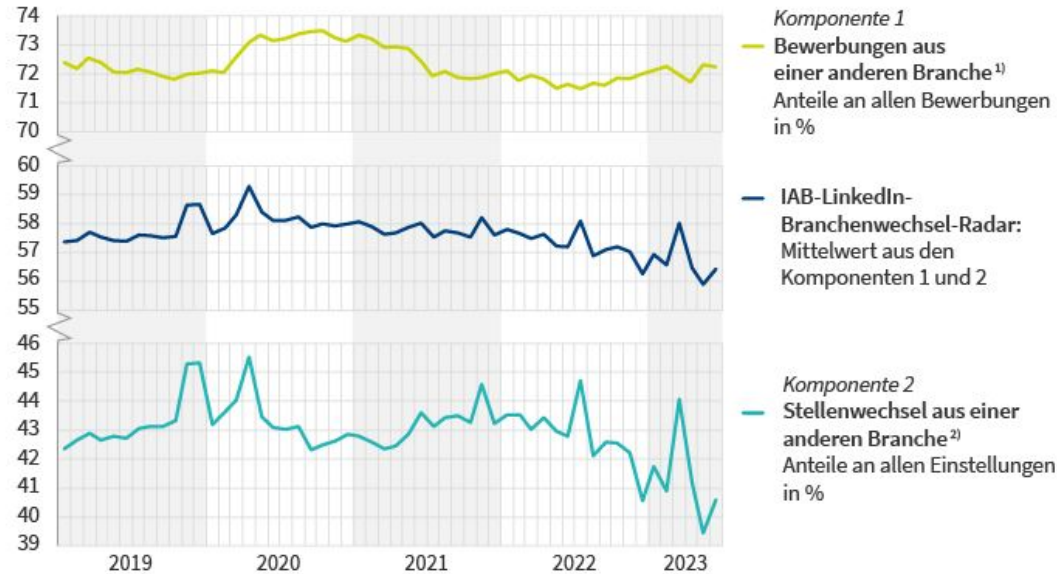
Interpretationshilfe: Hellere Gitterzellen weisen auf einen niedrigeren Gini-Koeffizienten – also niedrigere Lohnungleichheit – hin, dunklere auf einen höheren Gini-Koeffizienten. Flächen ohne Gitterzellen sind nicht oder zu schwach besiedelt und somit zensiert. Die zugrunde liegenden Karten zeigen schwach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

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

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

# Fragen beantworten mit (den richtigen) Daten

Abb: Das IAB-LinkedIn-Branchenwechsel-Radar



<sup>1)</sup> 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.

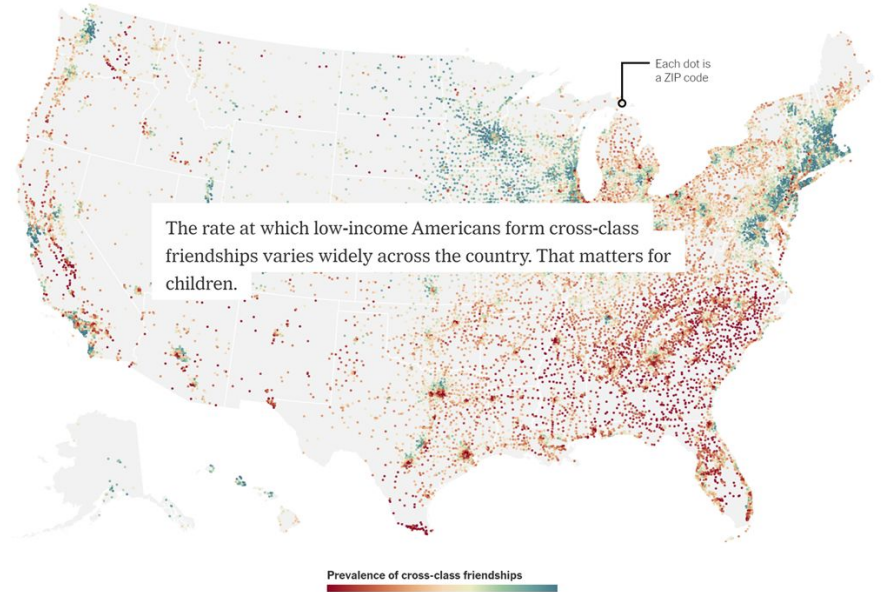
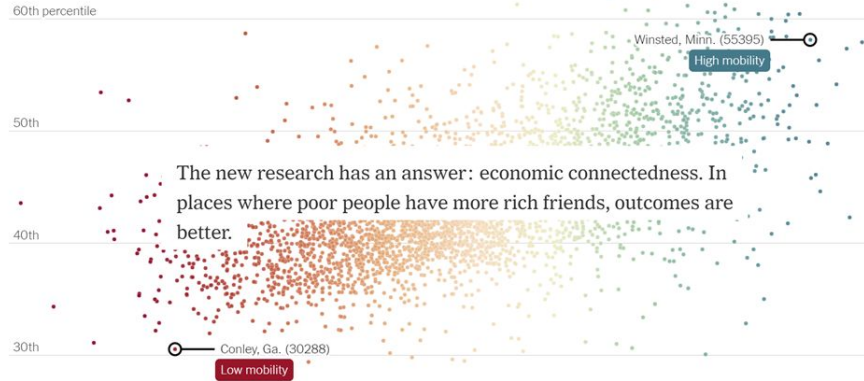
<sup>2)</sup> LinkedIn-Mitglieder, in deren Mitgliederprofil vor Antritt der neuen Stelle eine Stelle in einer anderen Branche eingetragen war.

Quelle: LinkedIn, eigene Berechnungen. © IAB

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

# Fragen beantworten mit kombinierten Daten

Average adult income rank  
for a poor child



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 "archivwürdig"?

Nutzen Sie für die Diskussion auch die [Gesetzlichen Vorgaben](#).



Was soll archiviert werden?

Was wird in 30 / 100 / .. Jahren wichtig sein?

Was bedeutet "archivarisch wertvoll"?

Könnte man empirisch lernen "Was ist von Interesse"?

Was bedeutet "berühmt"?



# 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

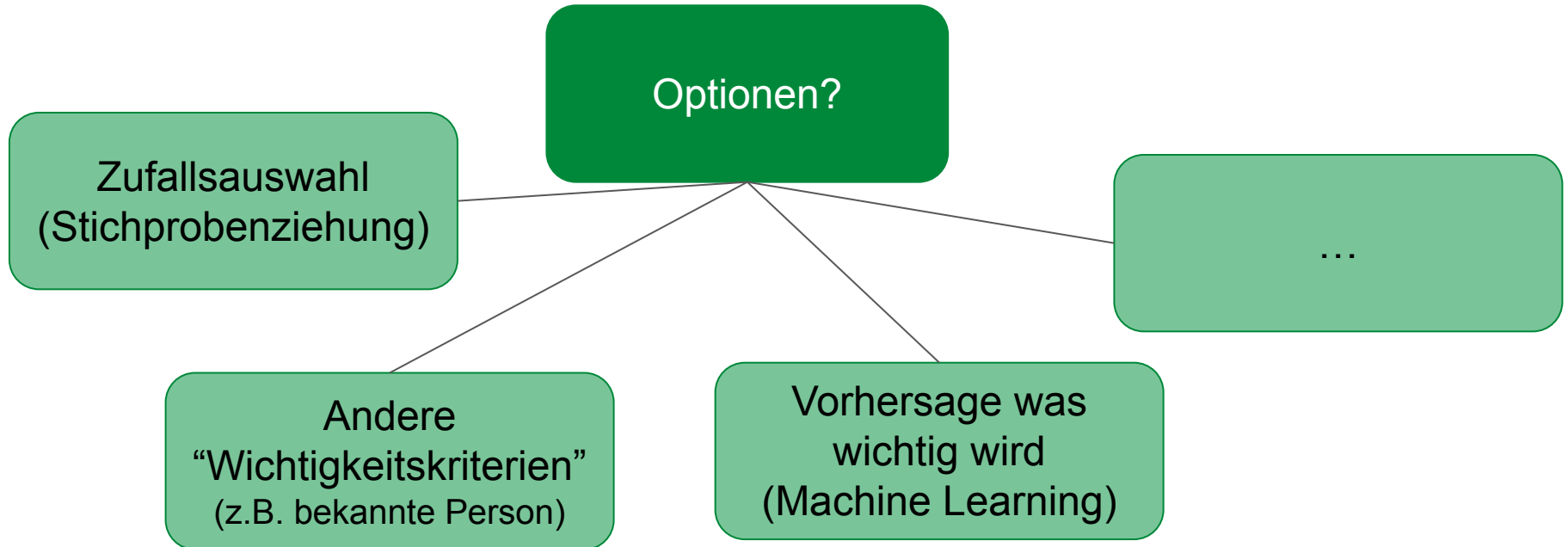
Christian Brück, Claudia Kalesse, Verena Christl, Sascha Bauer, Franziska Armbruster, Paul Manthey, Felix Henninger

Team 5 - Zoom

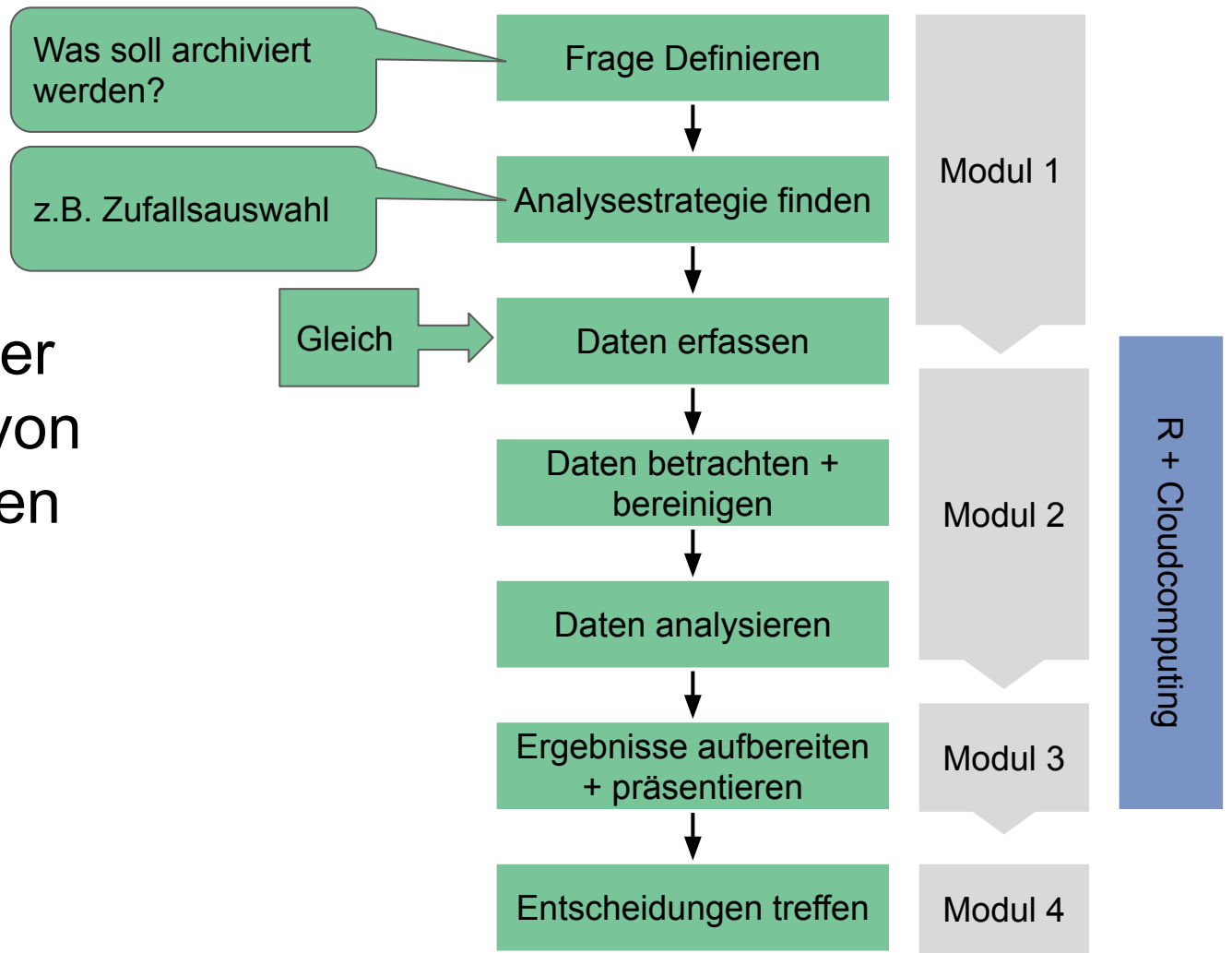
Gunther Friedrich, Ulrike Claudia Hofmann, Rhea Nguyen, Anita Bott, Dominik Mayer, Carolin Stimmelmayer, Marcel Neunhoeffer

# Daten nutzen, um eigene Probleme zu lösen.

Datenbasiert entscheiden welche Akten langfristig archiviert werden sollen.



# Vorgehen bei der Beantwortung von Fragen mit Daten



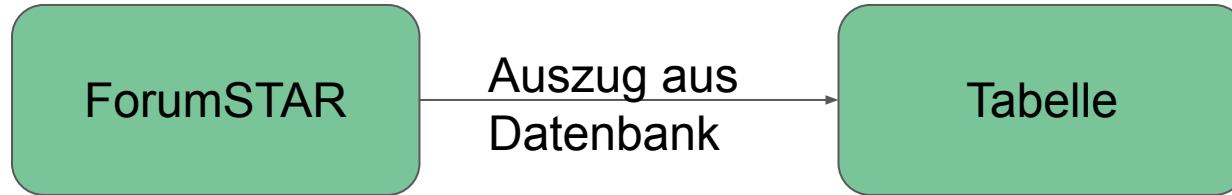


# Was sind Daten?



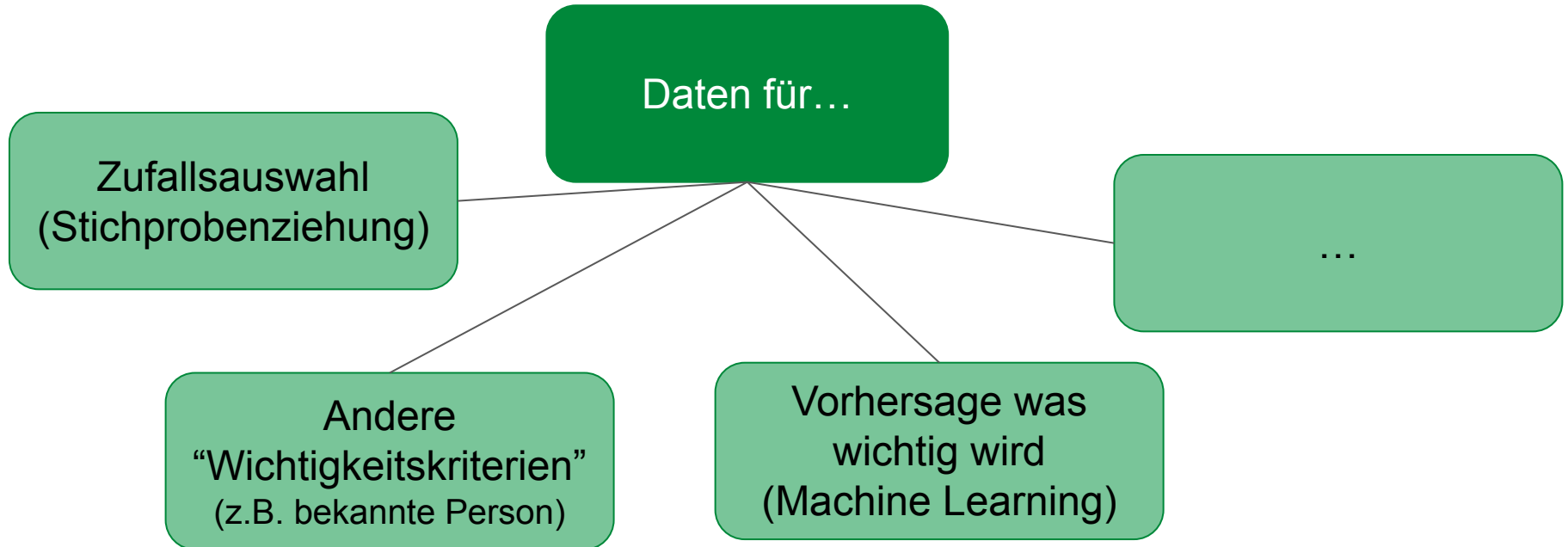
Mit welchen Arten von Daten beschäftigen Sie sich?

# Welche Daten gibt es in unserem Projekt?



FTCAM		<b>Daten betr. Scheidung</b>		örtliche Zuständigkeit ?	F_Sch_ROM II
Az: <b>1 F 1/15</b>		Wurde dieses Verfahren nach dem 20.06.2012 eingeleitet?		<input checked="" type="checkbox"/> ja <input type="checkbox"/> nein	
Heirat: 15.08.1990		Ehezeitbeginn: 01.08.1990 ? §		ZU SchAntrag: 18.01.2015	
vor dem Standesbeamten ? <input checked="" type="checkbox"/> Ja <input type="checkbox"/> Nein		Sind beide Ehegatten Iraner? <input type="checkbox"/> ja <input checked="" type="checkbox"/> nein		Ehezeitende: 31.12.2014 ?	
Standesamt: Rosenheim		Staat: (wenn Heiratsort im Ausland) Deutschland		?	
HeiratRegNr: 15/1990					
Hatten die Ehegatten zwischen Rechtshängigkeit und letzter mdl. Verhandlung zu irgendeinem Zeitpunkt - gemeinsam ihren gew. Aufenthaltsort in Deutschland?		<input checked="" type="checkbox"/> ja <input type="checkbox"/> nein		?	
Haben die Ehegatten eine Rechtswahl getroffen?		<input type="checkbox"/> ja <input checked="" type="checkbox"/> nein			
Hat es früher eine Trennungsentscheidung gegeben (Art. 9 ROM III-VO)?		<input type="checkbox"/> ja <input checked="" type="checkbox"/> nein		?	
Hatten beide Ehegatten bei Anrufung des Gerichts ihren gew. Aufenthaltsort in Deutschland?		<input checked="" type="checkbox"/> ja <input type="checkbox"/> nein			
Staatsangehörigkeit bei ZU des Scheidungsantrags (wegen VA gem. Art. 17 Abs. 3 EGBGB):		Ehemann: Türkei		Ehefrau: Türkei	
<input checked="" type="checkbox"/> außerdem - nicht "effektiv" -		<input type="checkbox"/> außerdem - nicht "effektiv" -		<input type="checkbox"/> außerdem - nicht "effektiv" -	
<input type="checkbox"/> Ehemann ist staatenlos		<input type="checkbox"/> Ehefrau ist staatenlos		<input type="checkbox"/> ja <input checked="" type="checkbox"/> nein	
War einer der Ehegatten bei ZU des Scheidungsantrags Flüchtling oder Asylberechtigter?					
Antragsteller/in <input checked="" type="checkbox"/> Ehefrau <input type="checkbox"/> Ehemann					
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	
Antragsgegner/in persönlich <input checked="" type="checkbox"/> stimmt zu <input type="checkbox"/> widerspricht		<input type="checkbox"/> widerspricht nicht		<input type="checkbox"/> äußert sich nicht	
Soll der Verfahrensteil Scheidung begründet werden?		<input checked="" type="checkbox"/> ja <input type="checkbox"/> nein			
Feststellungen des Gerichts zur Dauer des Getrenntlebens		<input type="checkbox"/> weniger als 1 Jahr (unzumutbare Härte)		<input checked="" type="checkbox"/> mindestens 1 Jahr	
		<a href="#">Generelle Vorgaben ändern</a>		<a href="#">Weiter</a>	
				F	

# 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



## Unsere Schnittstellen

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



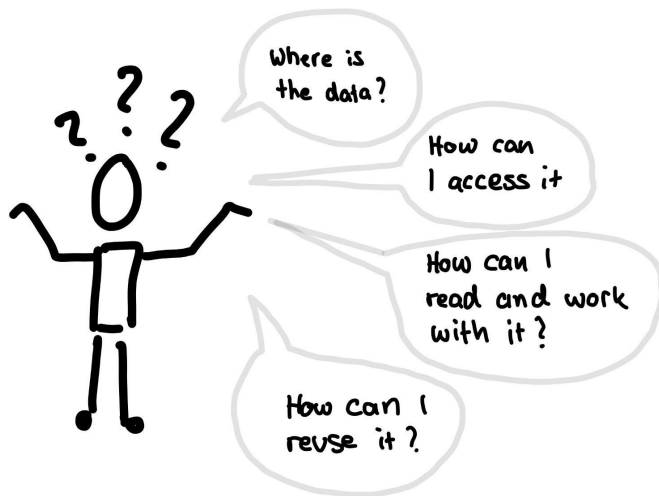
ACCESSIBLE



INTEROPERABLE



REUSABLE



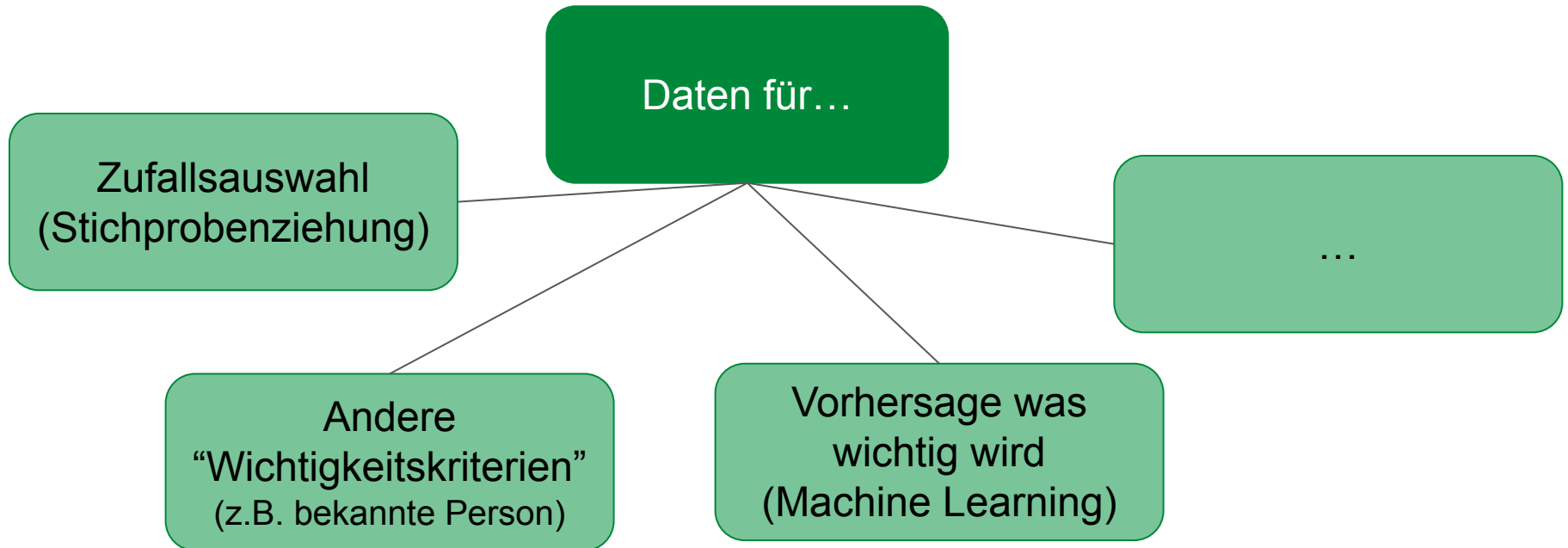
F A I R

≠

OPEN

---

# Woher könnten Daten kommen?



# Woher könnten Daten kommen?

Zufallsauswahl  
(Stichprobenziehung)

Daten für...



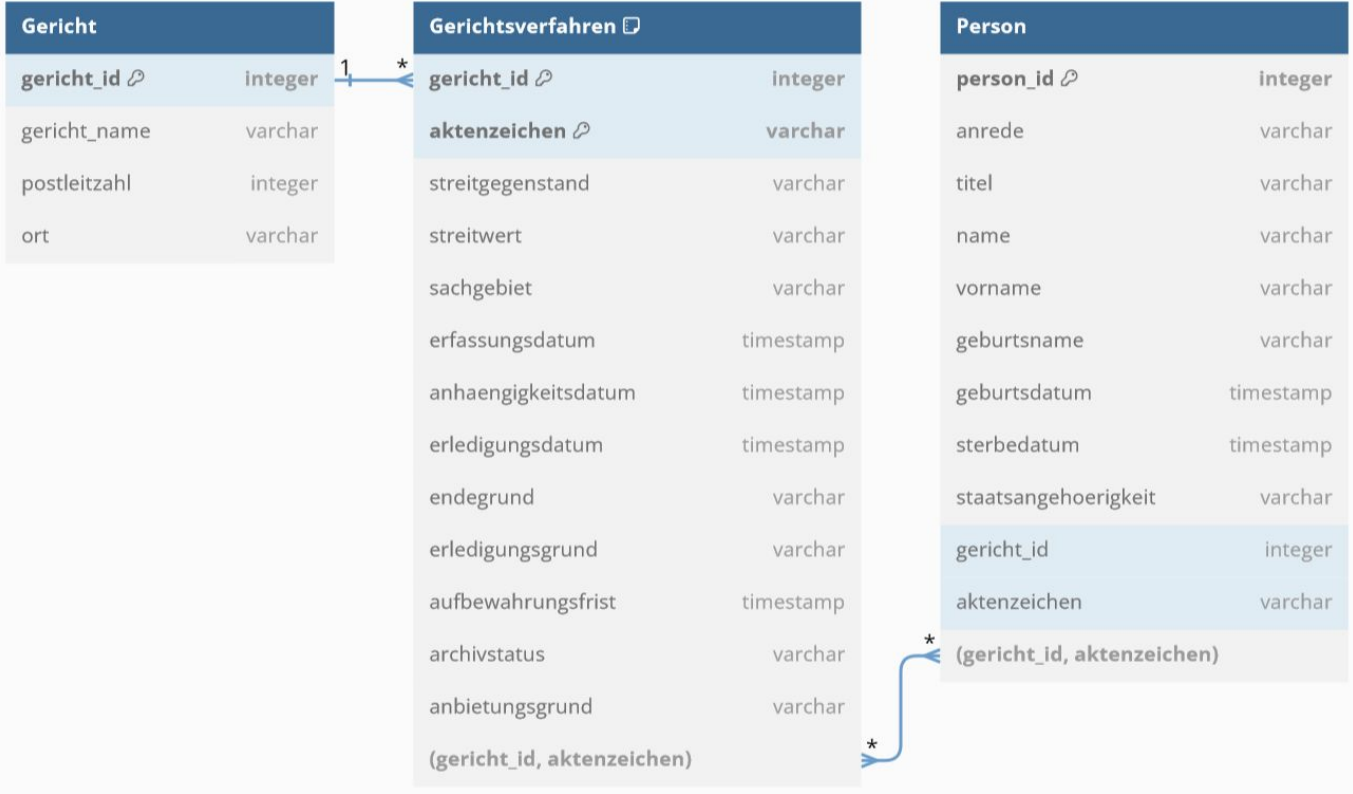
# Wie entstehen die verfügbaren Daten?



# 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, Streitgegenstand
2 FROM cases;
```

cases 1 x

SELECT Gericht, Aktenzeichen, Verfahrensstatus, Kurzrubrum, "Streitwert in EURO", Gesamtstreitgegenstand | Geben Sie einen SQL-Ausdruck ein, um die Ergebnisse zu filtern (verwenden Sie Strg+ Leertaste).

	Gericht	Aktenzeichen	Verfahrensstatus	Kurzrubrum	Streitwert	Gesamtstreitgegenstand
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: Personen

```
7 SELECT * from people;
```

people 1 x

SELECT \* from people | Geben Sie einen SQL-Ausdruck ein, um die Ergebnisse zu filtern (verwenden Sie Strg+ Leertaste).

	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	Beklagter	Frau
4	Amtsgericht Erlangen	1 C 100/18	Beklagter	Herr
5	Amtsgericht Erlangen	1 C 100/18	Beklagter	Herr
6	Amtsgericht Erlangen	1 C 100/18	Beklagter	Frau

# Tabelle: Schema

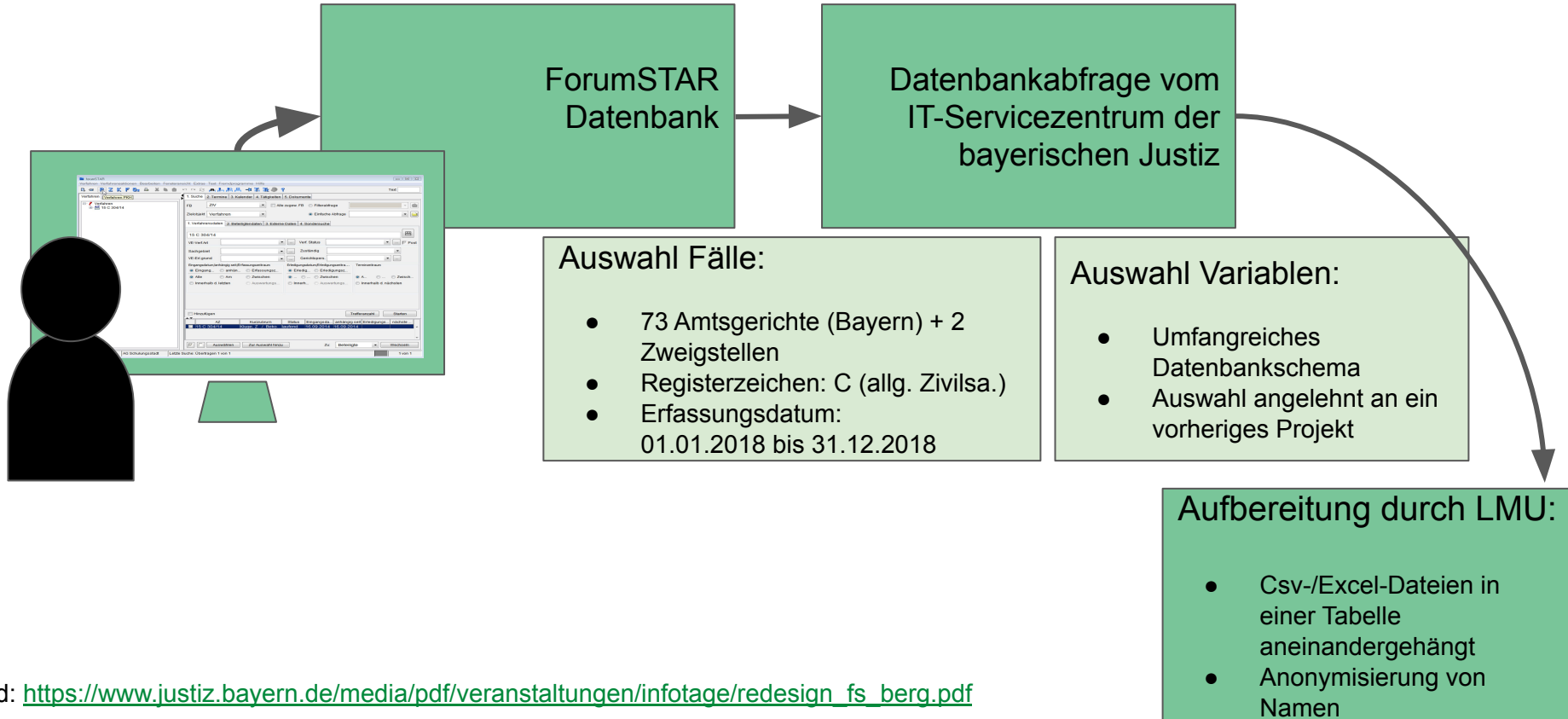
```
5 SELECT * FROM sqlite_schema;
```

sqlite\_master 1 x

SELECT \* FROM sqlite\_schema | Geben Sie einen SQL-Ausdruck ein, um die Ergebnisse zu filtern (verwenden Sie Strg+ Leertaste).

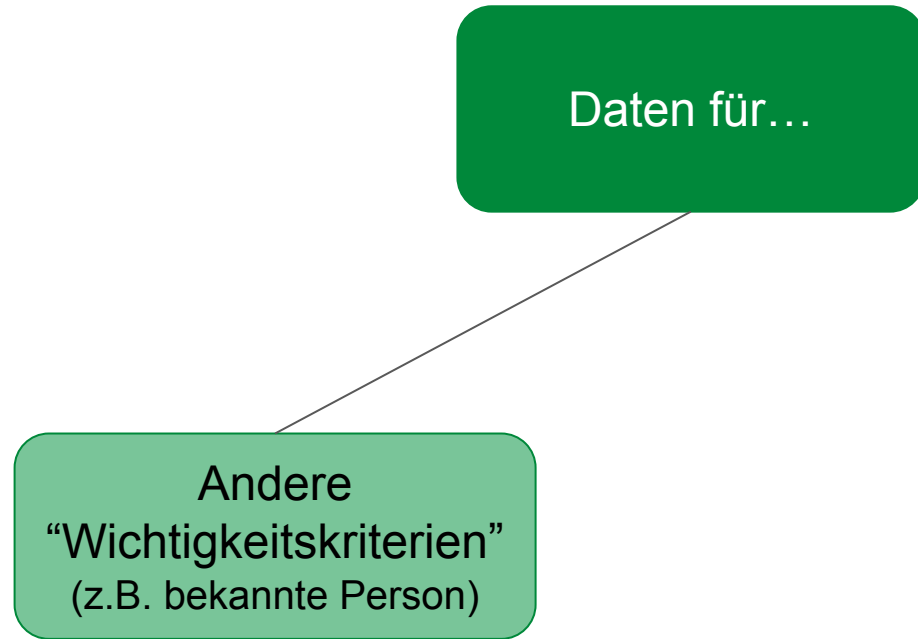
	type	name	tbl_name	rootpage	sql
1	table	sqlite_stat1	sqlite_stat1	121	CREATE TABLE sqlite_stat1(tbl,idx,stat)
2	table	sqlite_stat4	sqlite_stat4	122	CREATE TABLE sqlite_stat4(tbl,idx,neq,nlt,ndlt,sample)
3	table	cases	cases	2	CREATE TABLE `cases` (fl `Gericht` TEXT,fl `Aktenzeichen` TEXT,fl `Verfahrensstatus`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

# Wie entstehen die verfügbaren Daten?



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

# Welche externen Datenquellen können genutzt werden?





```

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

&lt;/&gt; Code

Herunterladen

Link

Search

person	personLabel	birthDate	description
<a href="#">wd:Q888991</a>	Lena Malkus	6. August 1993	deutsche Weitspringerin
<a href="#">wd:Q89097</a>	Carmen Klaschka	8. Januar 1987	deutsche Tennisspielerin
<a href="#">wd:Q89128</a>	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?

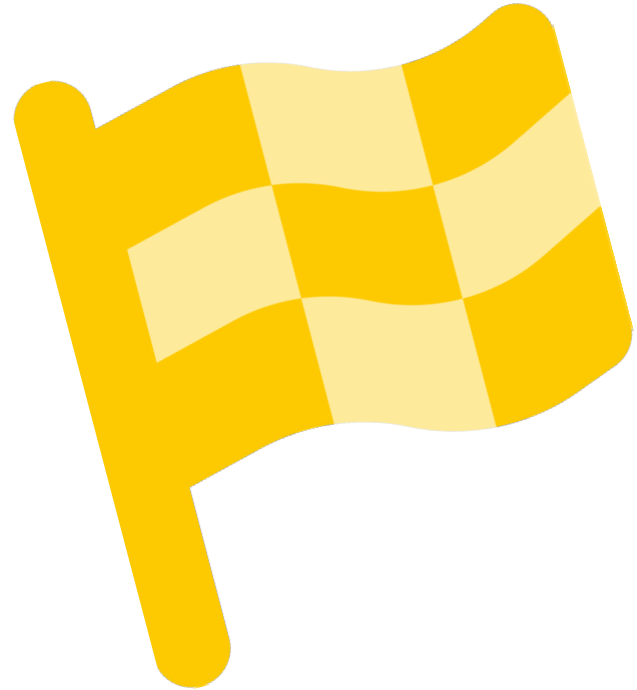




# Agenda

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



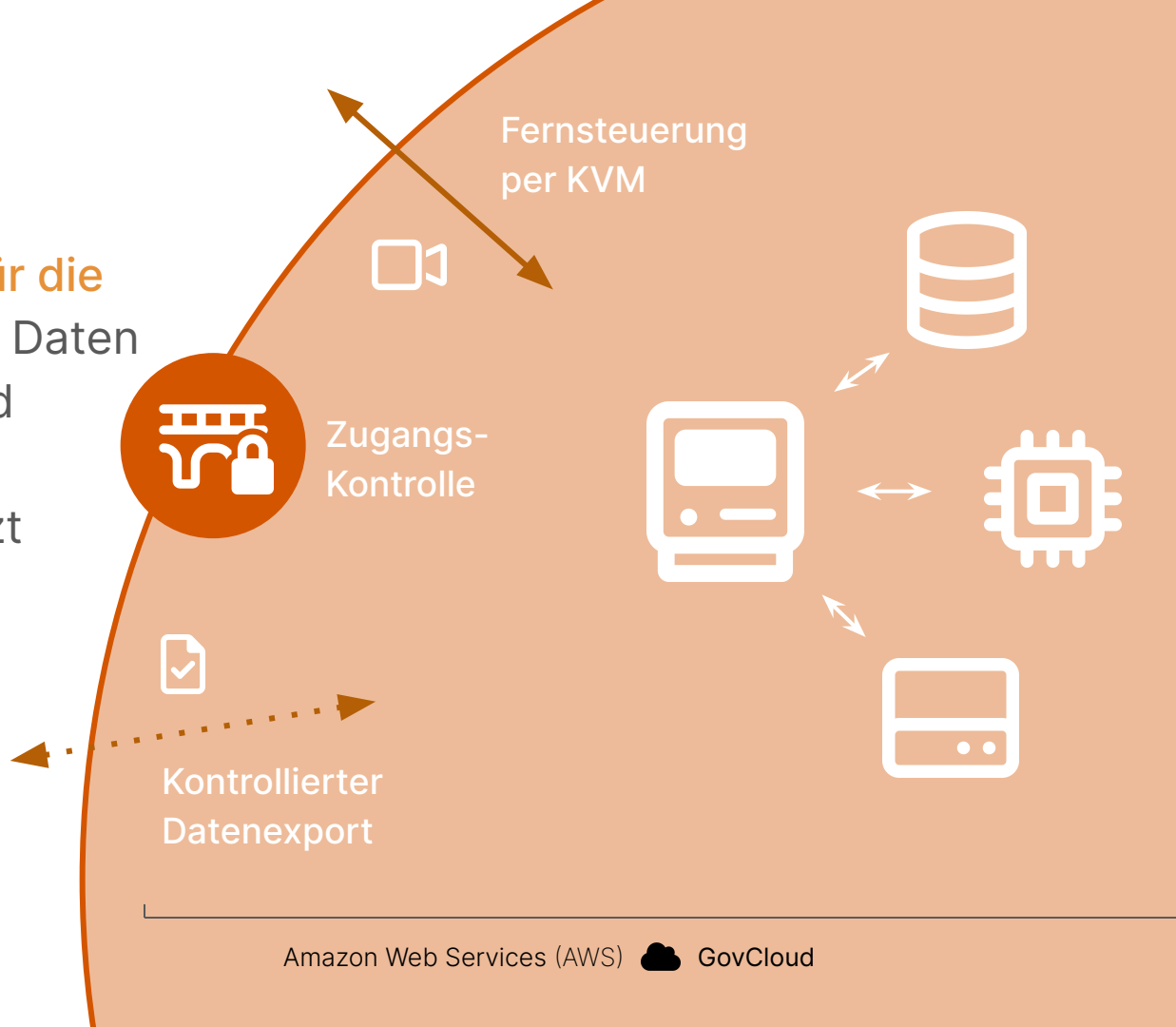
# Safe settings

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



# Safe settings

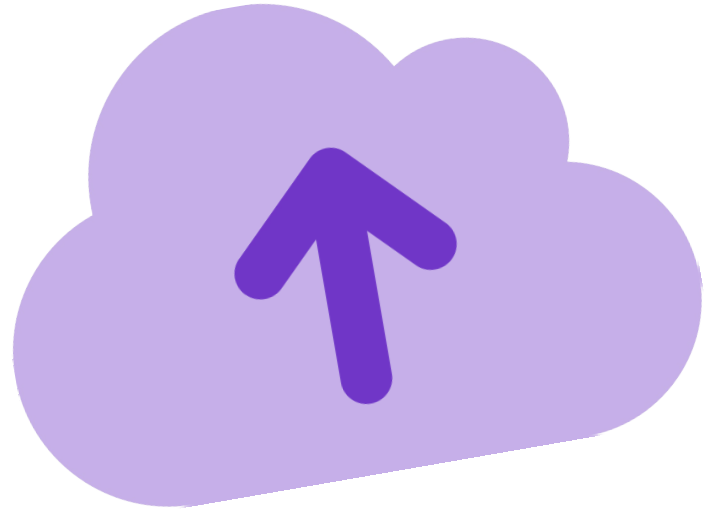
Geschützter Rahmen für die Datenverarbeitung, der Daten vor äußerem Zugriff und unintendierter Veröffentlichung schützt



# 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

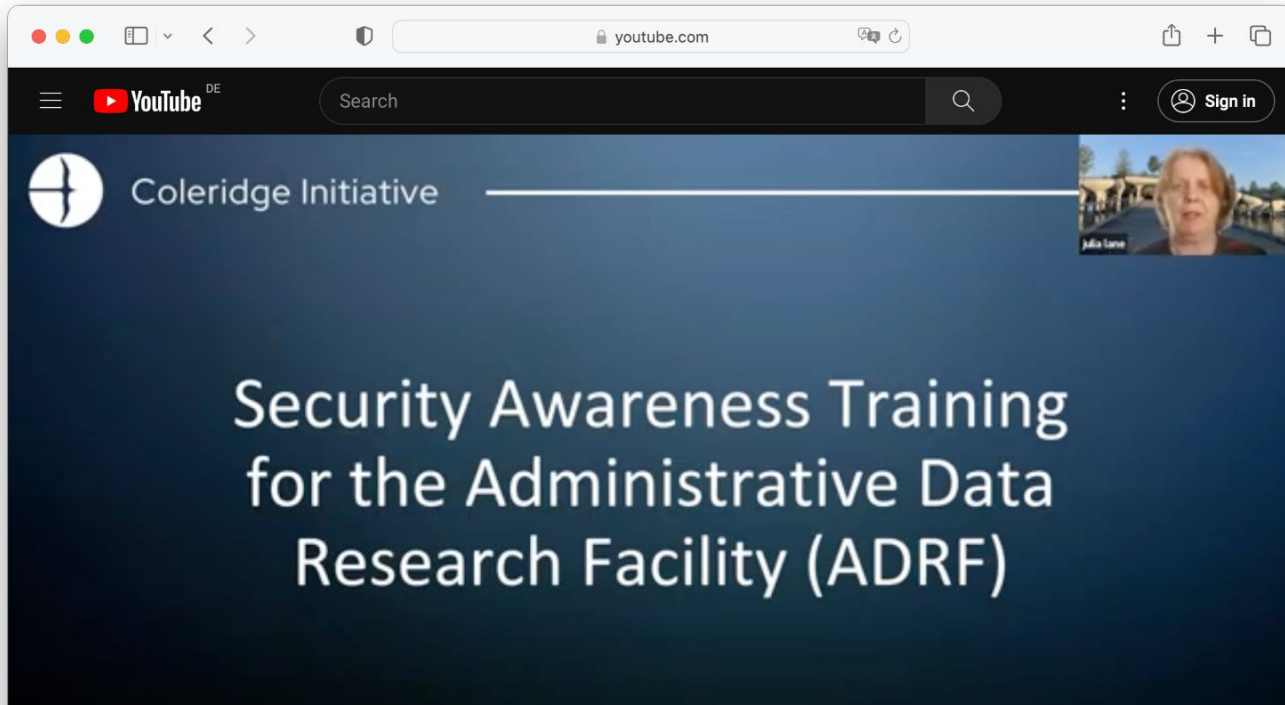
Bitte betrachten Sie den Zugang zu Daten als Privileg, und behandeln Sie sie als wären es Ihre eigenen

- 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>

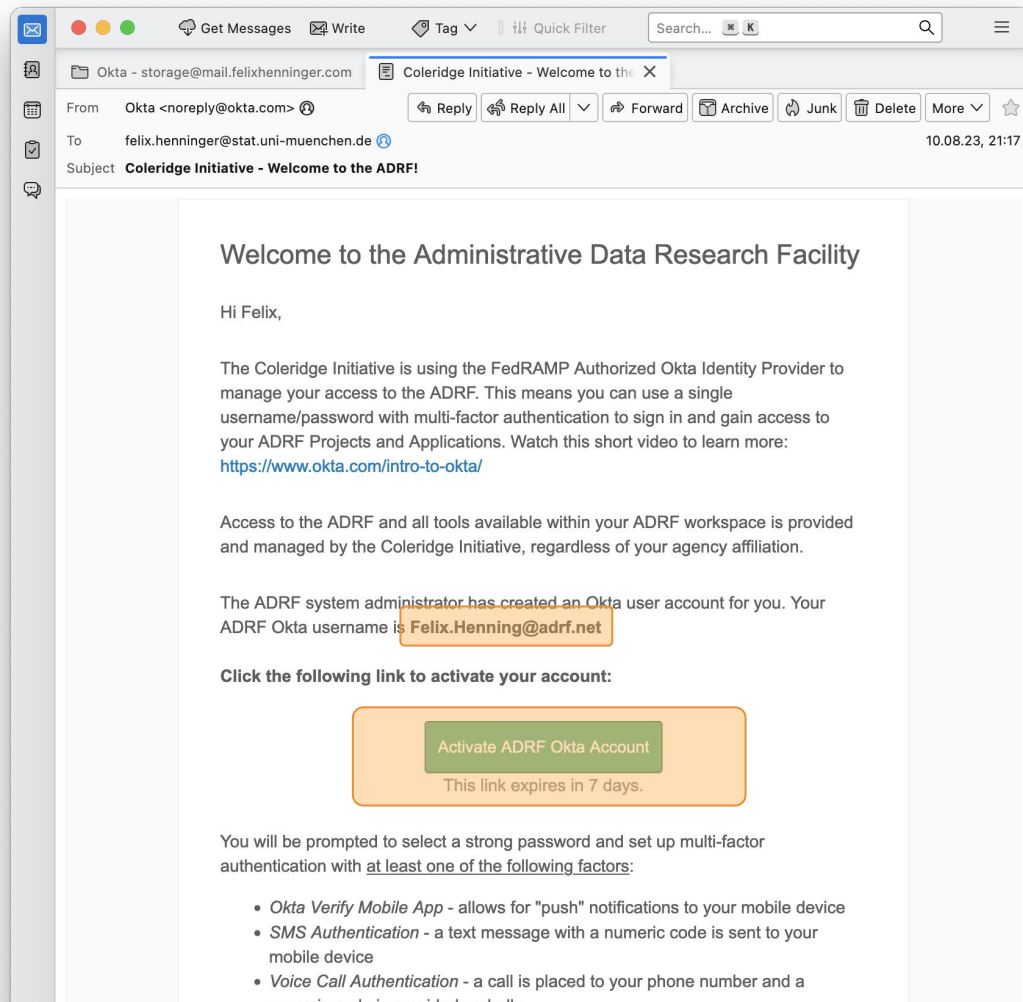


Demo



# Was bisher geschah ...

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



The screenshot shows an email interface with the following details:

- From:** Okta <noreply@okta.com>
- To:** felix.henninger@stat.uni-muenchen.de
- Subject:** Coleridge Initiative - Welcome to the ADRF!
- Date:** 10.08.23, 21:17

The email content includes:

### 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@adrf.net**

**Click the following link to activate your account:**

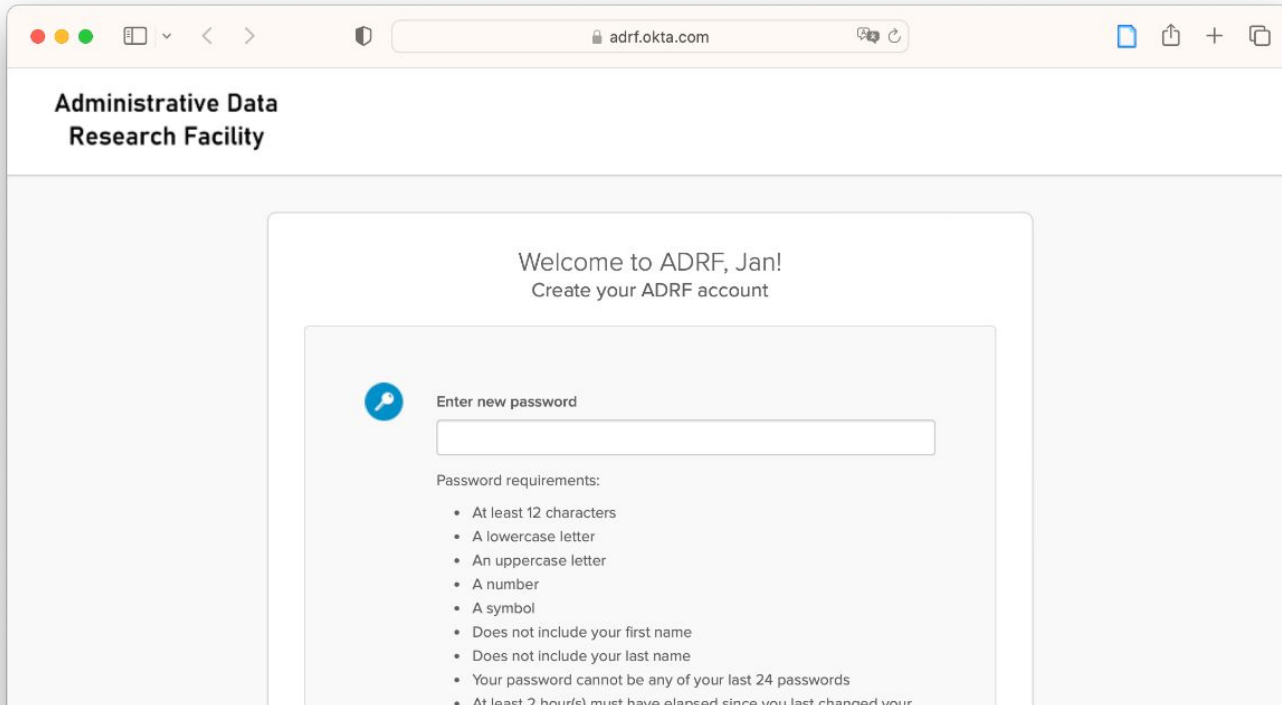
[Activate ADRF Okta Account](#)  
This link expires in 7 days.

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

# Registrierung

Auf Okta können Sie einen Zugang anlegen




The screenshot shows a web browser window with the URL `adrf.okta.com`. The page title is "Administrative Data Research Facility". The main content area displays a welcome message: "Welcome to ADRF, Jan! Create your ADRF account". Below this, there is a section titled "Enter new password" with a blue lock icon and a text input field. Underneath the input field, the "Password requirements:" are listed as follows:

- At least 12 characters
- A lowercase letter
- An uppercase letter
- A number
- A symbol
- Does not include your first name
- Does not include your last name
- Your password cannot be any of your last 24 passwords
- At least 2 hour(s) must have elapsed since you last changed your

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.

Optionen zur  
Passwort-  
wieder-  
herstellung



## Sicherheits- Bild

(wird beim  
login  
angezeigt)

+ Add Phone Number



Click a picture to choose a security image

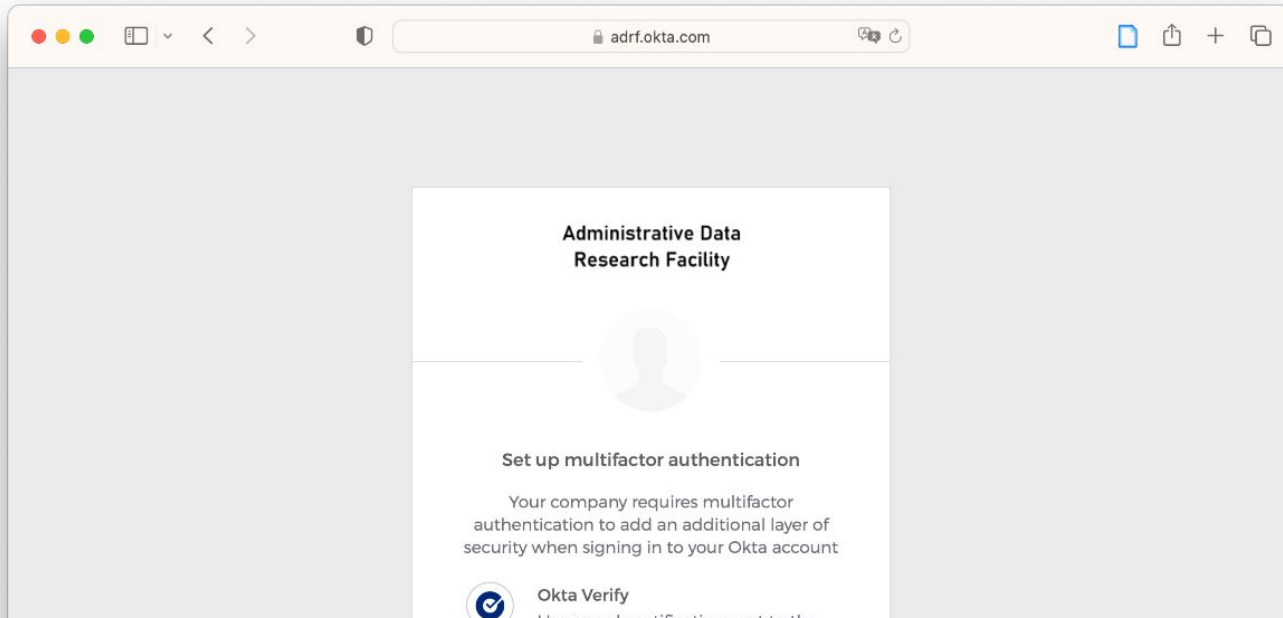
Your security image gives you additional assurance that you are logging into Okta, and not a fraudulent website.



Create My Account

# Zwei-Faktor-Authentifizierung

Jeder Login wird über einen unabhängigen Kanal bestätigt



## 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

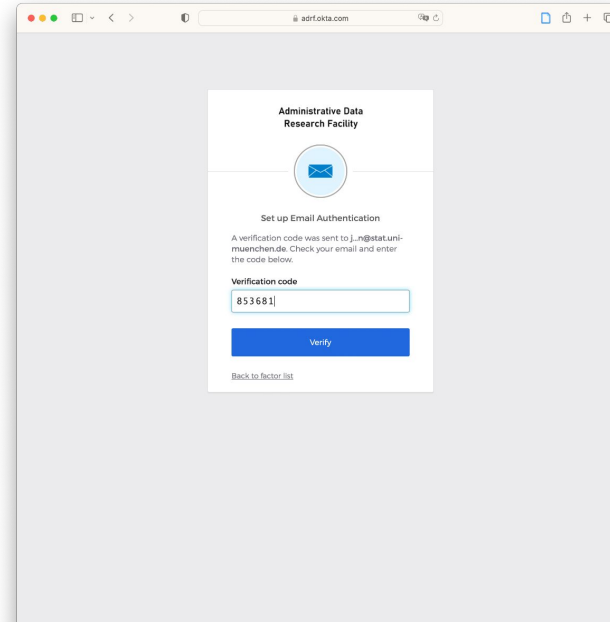
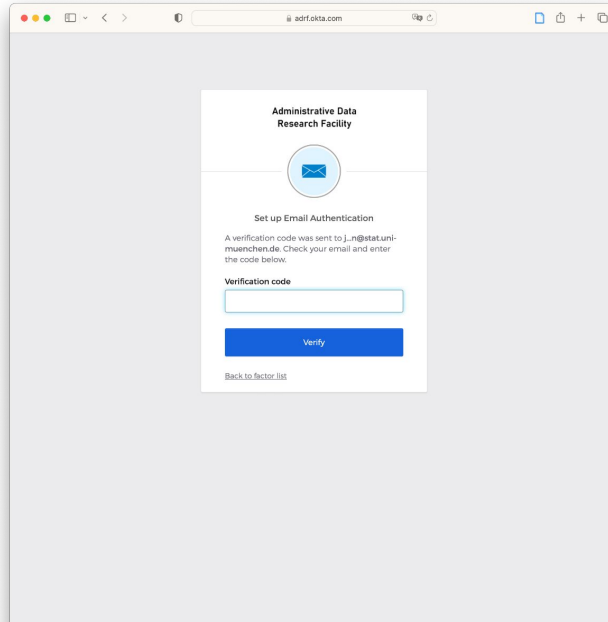
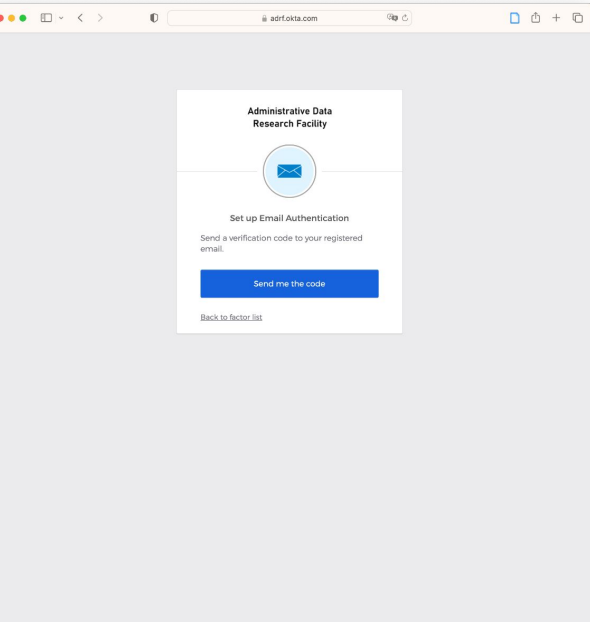
Enter a verification code sent to your email.

[Setup](#)

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

# Login

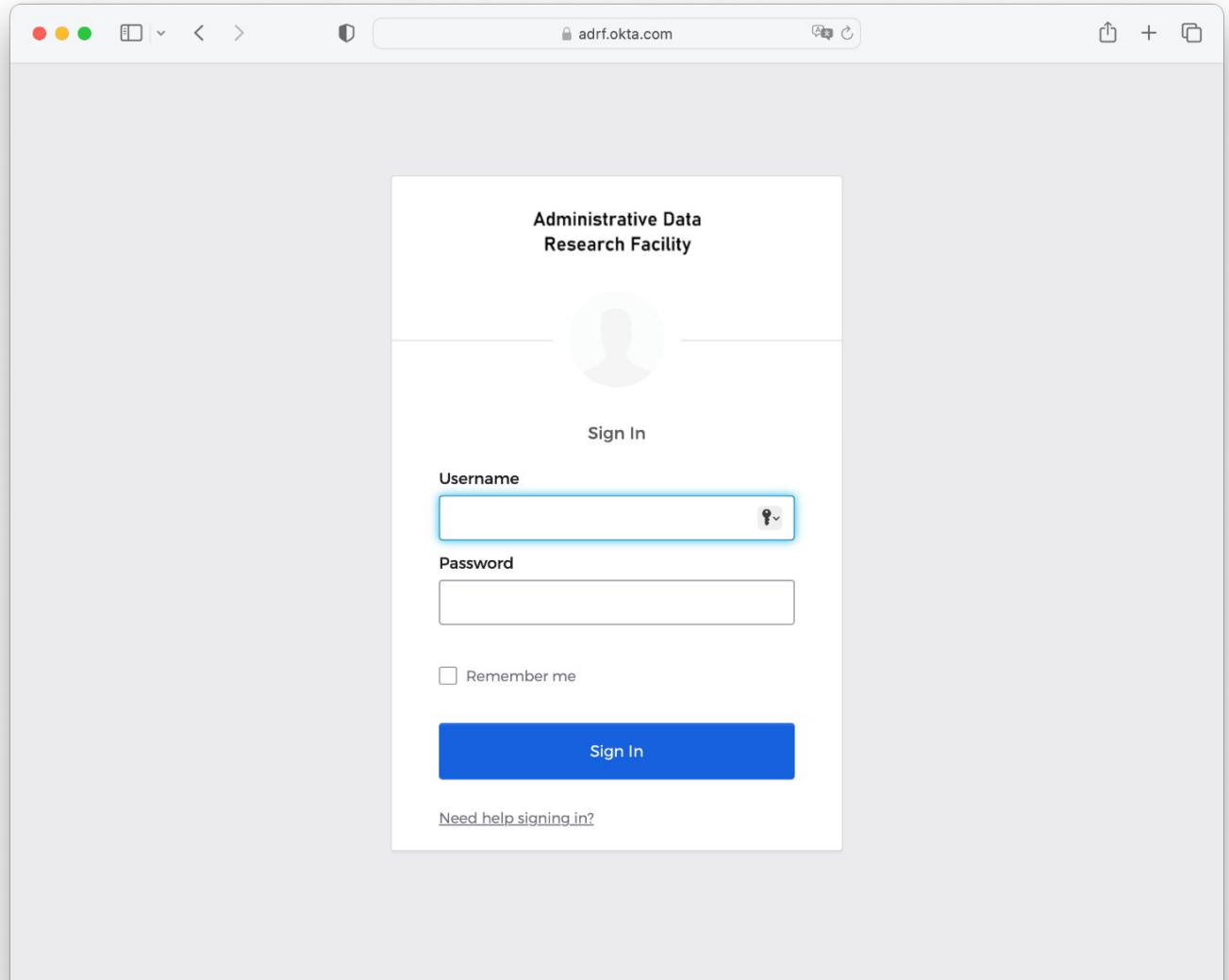
Über Ihren gewählten zweiten Faktor erhalten Sie einen Freischaltungscode



# Login

Zugang erhalten Sie  
immer über die  
Adresse

<https://adrf.okta.com>



The screenshot shows a web browser window with the address bar displaying "adrf.okta.com". The page content is centered and features the following elements:

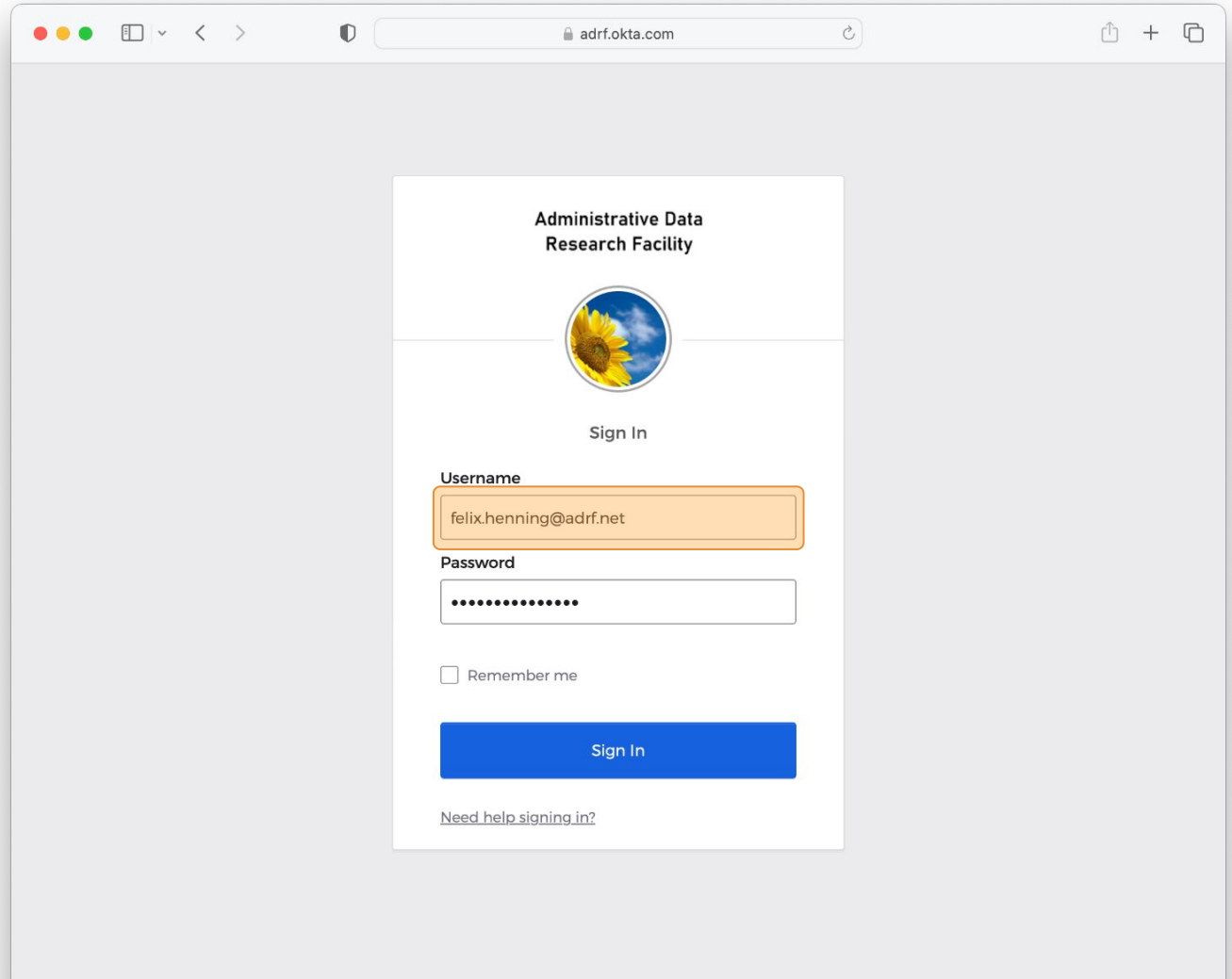
- Administrative Data Research Facility** (Title)
- A circular placeholder for a user profile picture.
- Sign In** (Text)
- Username** label above a text input field with a visibility toggle icon.
- Password** label above a text input field.
- Remember me (checkbox and text)
- Sign In** (Blue button)
- [Need help signing in?](#) (Link)

# Login


Zugang erhalten Sie  
immer über die  
Adresse

<https://adrf.okta.com>

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



Administrative Data  
Research Facility



Sign In

Username  
felix.henning@adrf.net

Password  
.....

Remember me

Sign In

[Need help signing in?](#)

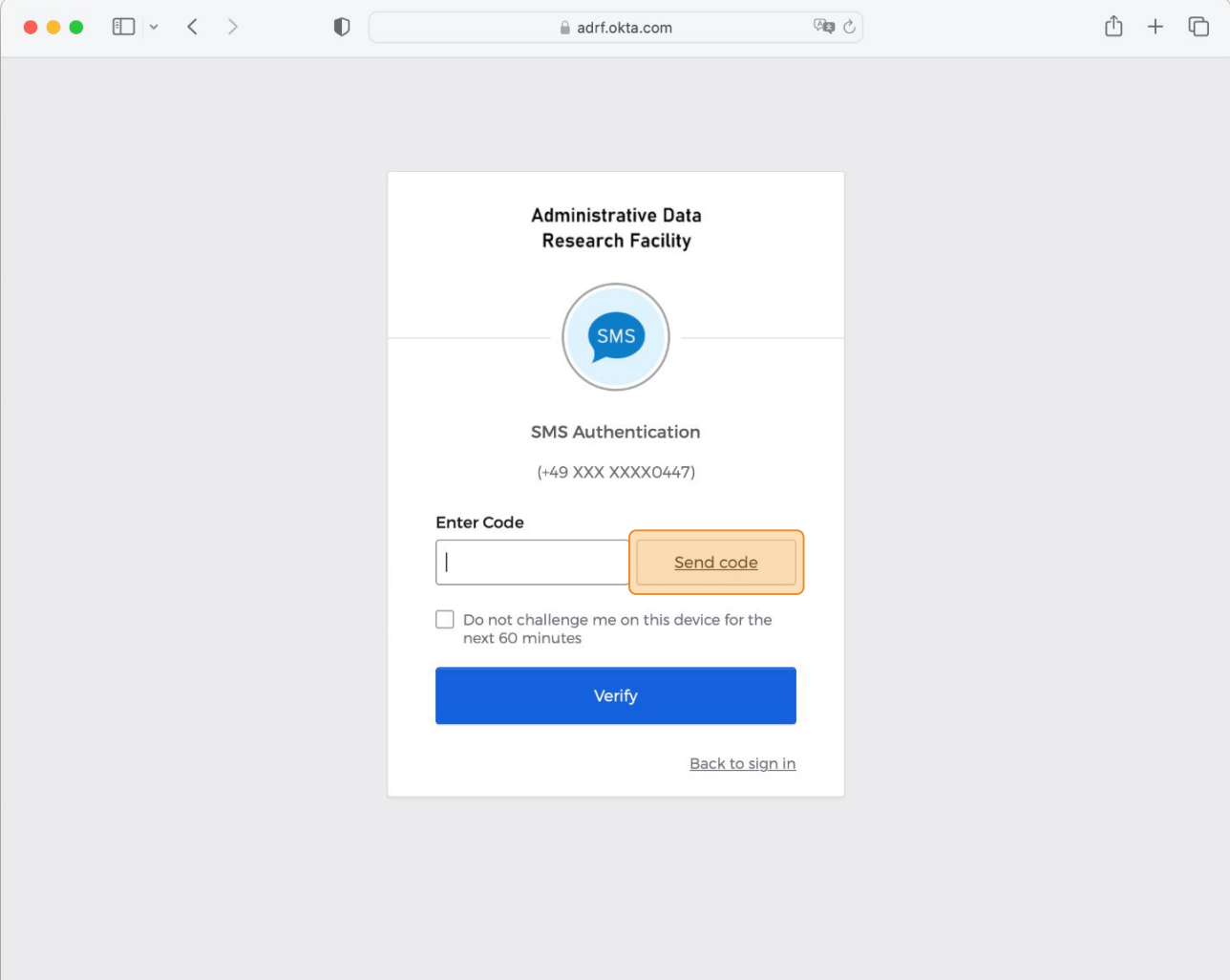
# Login

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.



The screenshot shows a web browser window with the URL [adrf.okta.com](https://adrf.okta.com). The page content is as follows:

**Administrative Data Research Facility**



**SMS Authentication**  
(+49 XXX XXXX0447)

**Enter Code**

Do not challenge me on this device for the next 60 minutes

[Back to sign in](#)

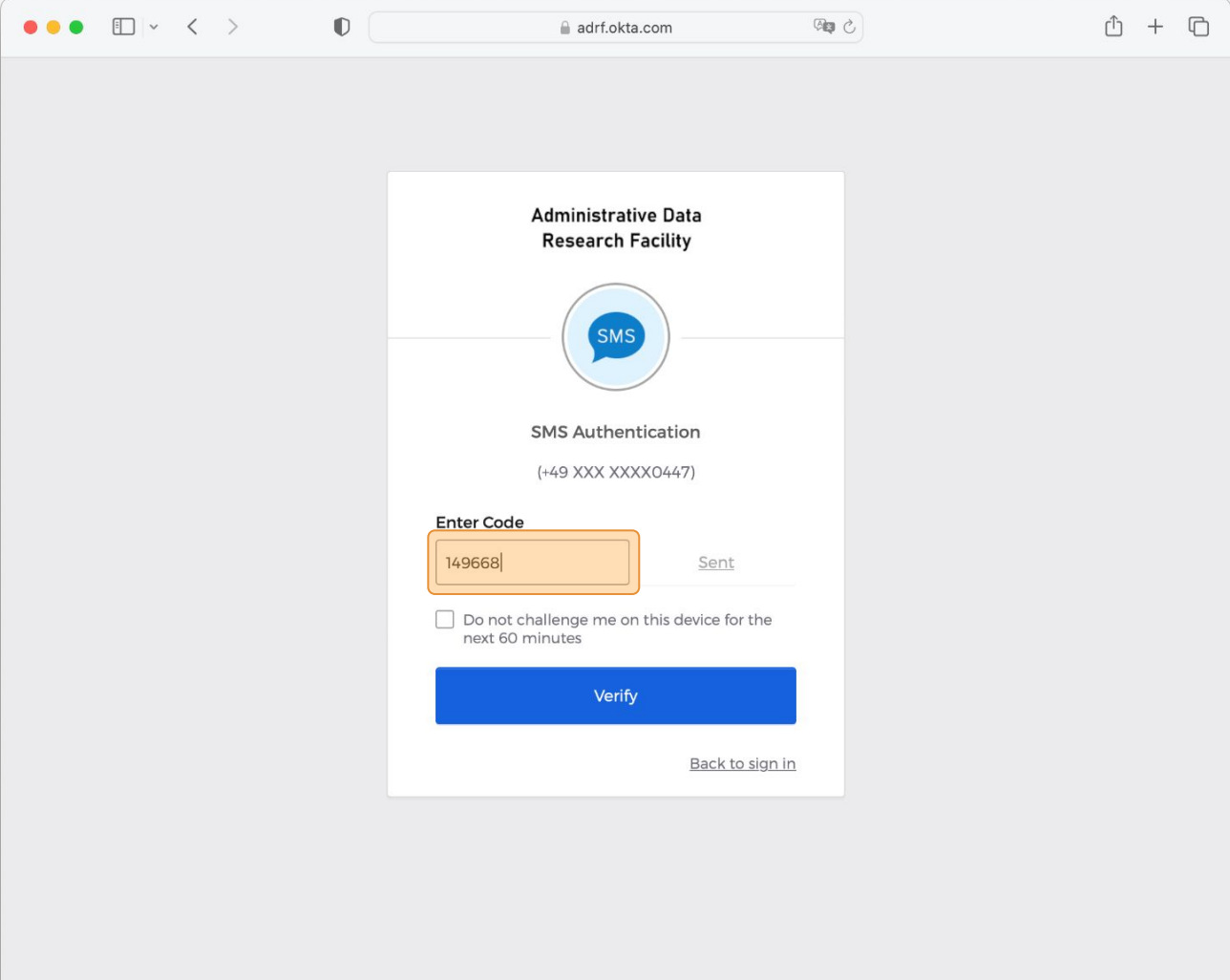
# Login

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.



The screenshot shows a web browser window with the URL [adrf.okta.com](https://adrf.okta.com). The page content is as follows:

**Administrative Data Research Facility**



**SMS Authentication**  
(+49 XXX XXXX0447)

**Enter Code**

[Sent](#)

Do not challenge me on this device for the next 60 minutes

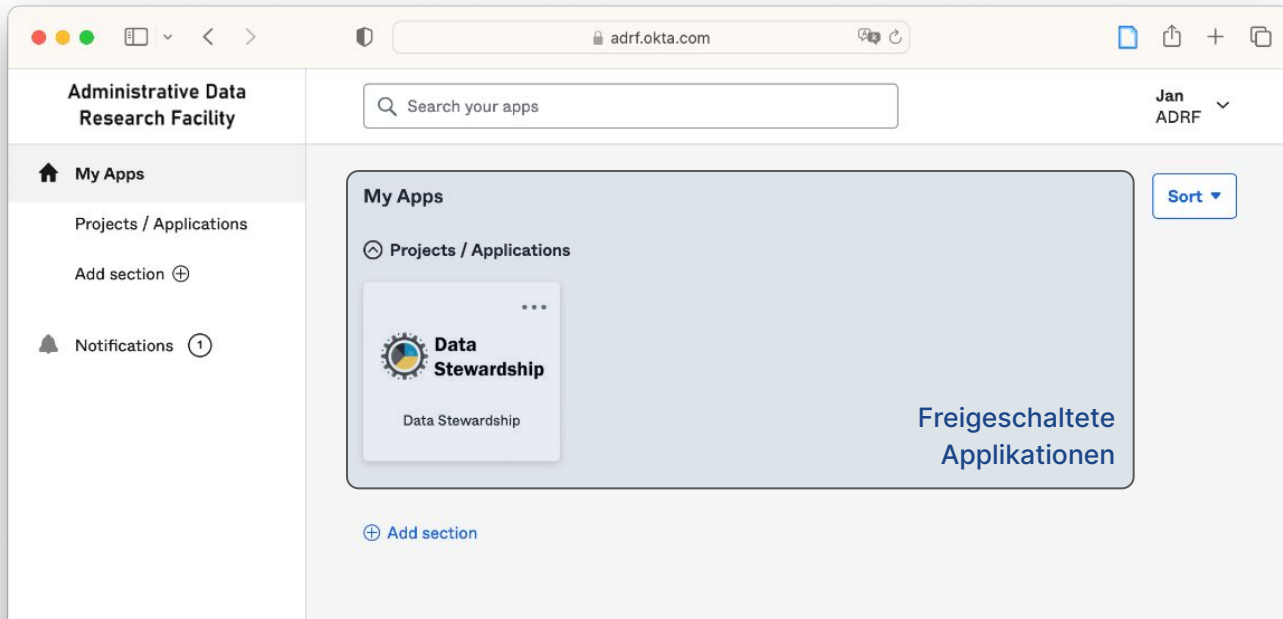
[Verify](#)

[Back to sign in](#)



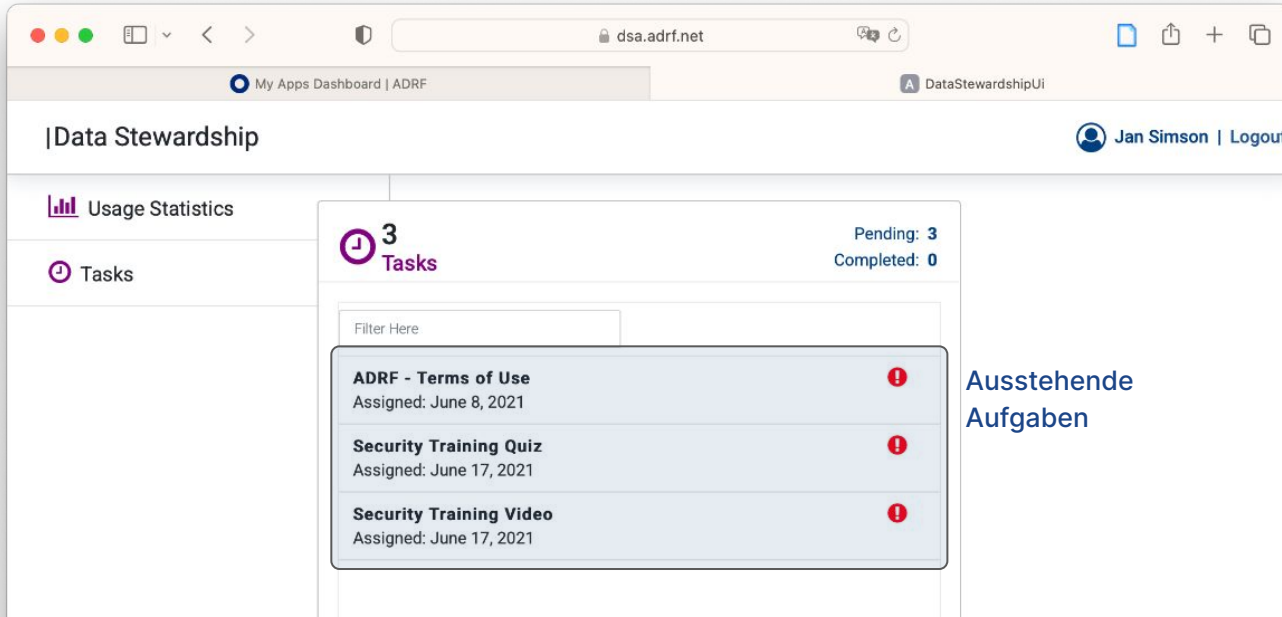
# Applikationen

Nach dem Login stehen Ihnen die Applikationen des ADRF zur Verfügung



# Data Stewardship-Modul

Prüft die Voraussetzungen für den Zugang zur Datenplattform



The screenshot shows a web browser window with the URL `dsa.adrf.net`. The page title is "Data Stewardship" and the user is logged in as "Jan Simson". The main content area displays a "Tasks" section with a filter input and a list of three pending tasks, each marked with a red exclamation point icon. The tasks are:

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

Summary statistics for the tasks are shown at the top right of the list: Pending: 3, Completed: 0.

A blue text label "Ausstehende Aufgaben" is positioned to the right of the task list.

# Zugangsvoraussetzungen

## Unterzeichnung der Nutzungsbedingungen

The screenshot shows a DocuSign envelope titled "Administrative Data Research Facility (ADRF) Terms of Use". The document text includes:

**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](mailto: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 in 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 over 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%."

ADRF - Terms of Use (June 2021).pdf 1 of 2

## Sicherheitsunterweisung

The screenshot shows the Data Stewardship dashboard with a "Security Quiz Video" overlay. The dashboard displays "Usage Statistics" and "Tasks" (3 pending, 0 completed). The video overlay features the title "Security Awareness Training for the Administrative Data Research Facility (ADRF)" and a YouTube link: <https://coleridgeinitiative.org/adrf/>. The video player includes "Mark Complete" and "Close" buttons.

## Quiz zur Sicherheit

The screenshot shows the "Security Training Quiz" interface. It asks the user to complete 5 of 6 questions correctly to pass. The questions are:

- It's OK to share data from the ADRF with my colleagues if I swear them to secrecy.  
 Yes  
 No
- It's OK to take a screenshot of a graph from the ADRF and put it in a PowerPoint.  
 Yes  
 No
- It's OK to write down summary information from a table in the ADRF and use it for a presentation.  
 Yes  
 No
- If someone else on the same project doesn't have their login information, it's OK to share my login information with them.  
 Yes  
 No
- 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.
- 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 submission. 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

Administrative Data Research Facility

Search your apps

Jan ADRF

My Apps

Projects / Applications

Add section

Notifications 1

My Apps

Sort

Projects / Applications

Data Stewardship

ADRF Management Portal

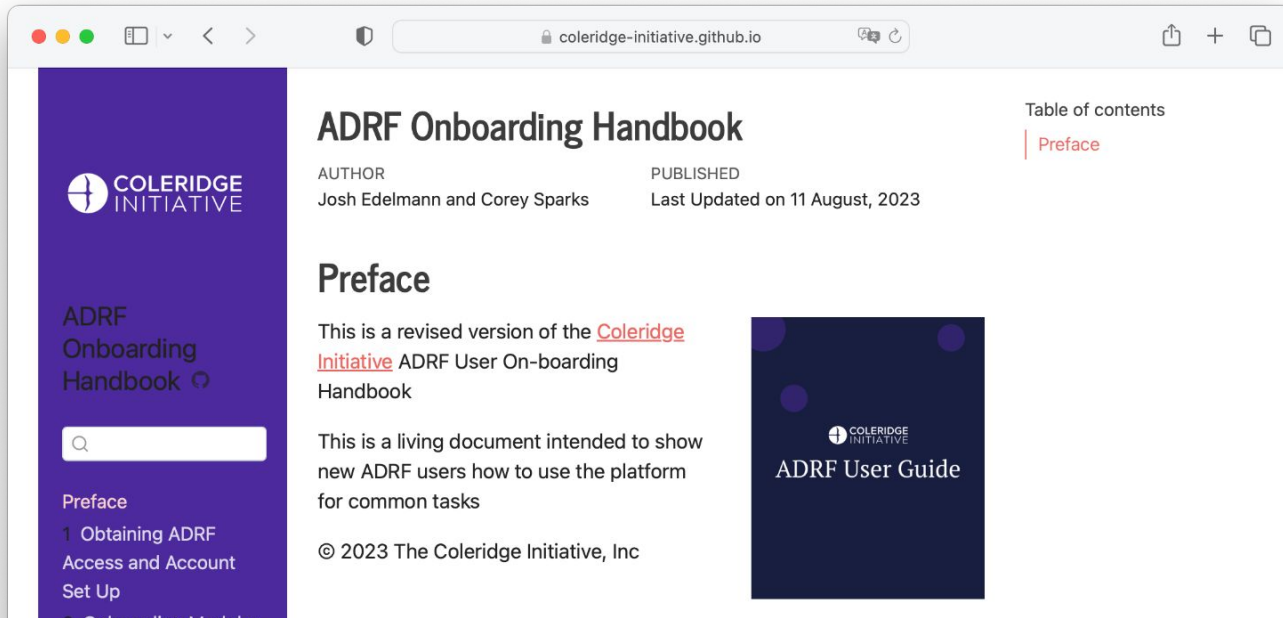
ADA Bayern

Zugang zur Arbeits-Umgebung

Add section

# Weitere Informationen ...

.. finden Sie im 'onboarding handbook' unter  
[https://coleridge-initiative.github.io/adrf\\_onboarding\\_handbook/](https://coleridge-initiative.github.io/adrf_onboarding_handbook/)



The screenshot shows a web browser window displaying the 'ADRF Onboarding Handbook' page. The browser's address bar shows the URL 'coleridge-initiative.github.io'. The page features a purple sidebar on the left with the Coleridge Initiative logo and a search bar. The main content area has a white background with a purple header for the handbook title. Below the title, it lists the authors (Josh Edelmann and Corey Sparks) and the publication date (Last Updated on 11 August, 2023). The 'Preface' section is highlighted, containing text about the handbook being a revised version of the Coleridge Initiative ADRF User On-boarding Handbook and a living document for new users. A 'Table of contents' link is visible in the top right, with 'Preface' selected. A dark blue box on the right side of the page contains the Coleridge Initiative logo and the text 'ADRF User Guide'.

ADRF Onboarding Handbook

Table of contents  
[Preface](#)

**ADRF Onboarding Handbook**

AUTHOR  
Josh Edelmann and Corey Sparks

PUBLISHED  
Last Updated on 11 August, 2023

## Preface

This is a revised version of the [Coleridge Initiative](#) ADRF User On-boarding Handbook

This is a living document intended to show new ADRF users how to use the platform for common tasks

© 2023 The Coleridge Initiative, Inc

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-Plattform schalten Sie danach (bis morgen) frei

