

Tag 3: Auf dem Weg zur datenbewussten Organisation

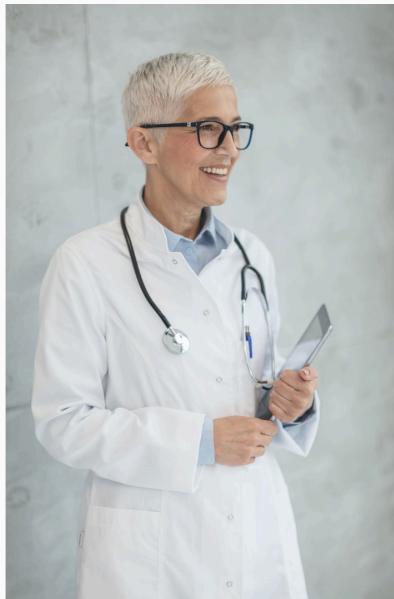
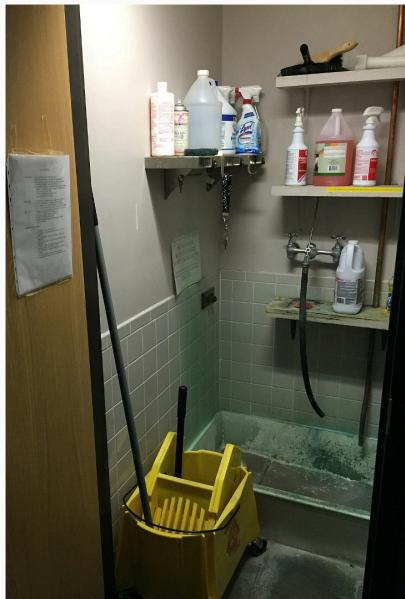
Session 9: Datenarchivierung und Veröffentlichung

Sebastian Ramirez-Ruiz
Hertie School

Was die Leute denken, wie die Arbeit mit Daten aussieht...

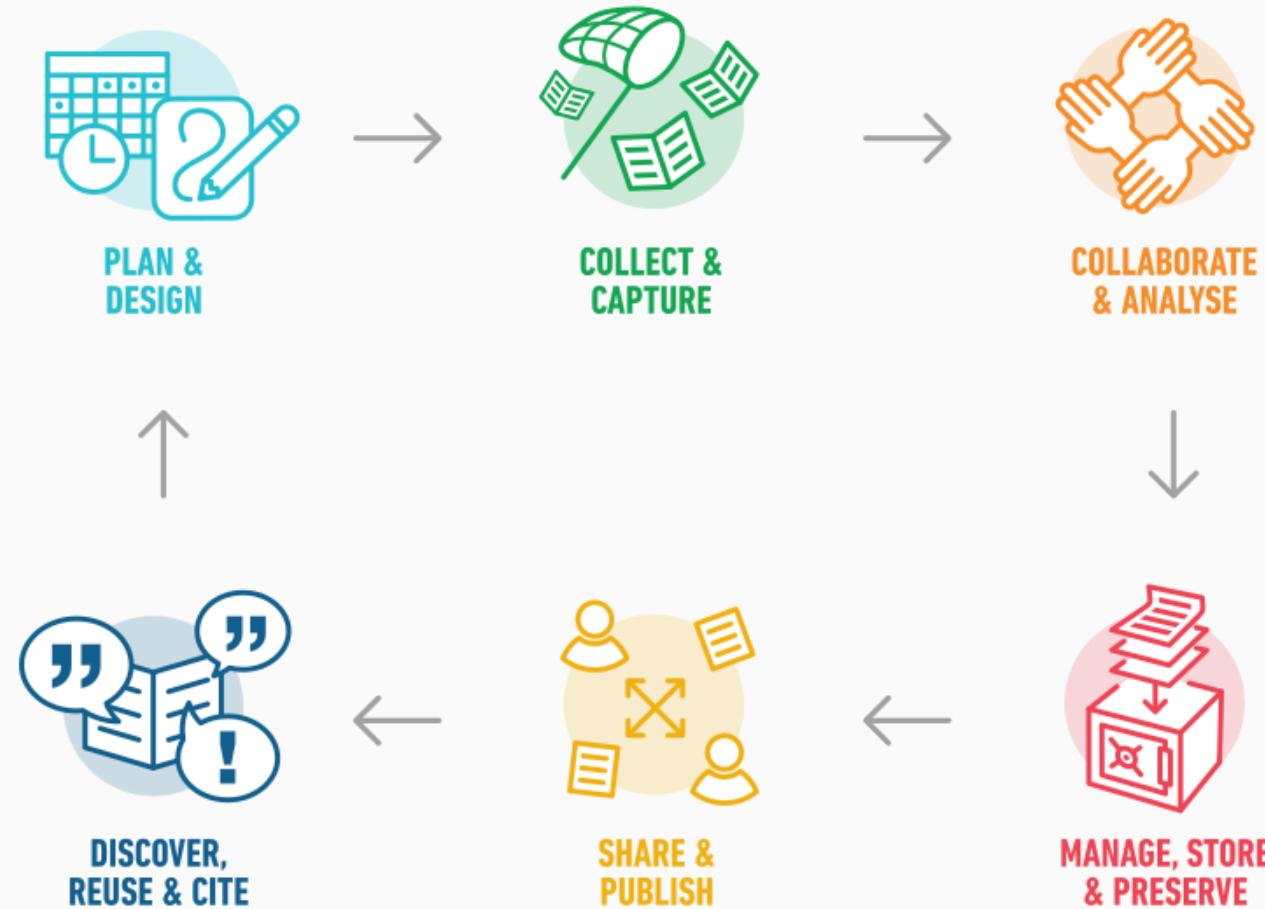


Wie es wirklich ist...



You have to wear many hats...

Research Data Management (RDM) Lifecycle



1. Datenmanagement und ein Datenmanagementplan (DMP)
2. Schutz
3. Organisieren, Dokumentieren, Verarbeiten und Speichern

Research Data Management (RDM) Lifecycle



Heben Sie die Hand, wenn Sie:

- den Begriff Datenmanagementplan (DMP) schon einmal gehört haben
- definieren können, was ein Datenmanagementplan beinhaltet
- schon einmal einen Datenmanagementplan für ein Projekt erstellt haben
- Datenerhebungsmethoden für Forschung oder politische Arbeit geplant haben
- Strategien zur Gewährleistung der Datenintegrität und -sicherheit umgesetzt haben
- mit den FAIR-Prinzipien (Findability, Accessibility, Interoperability und Reuse) vertraut sind
- Sie haben Praktiken für offene Daten in Ihre Arbeit integriert
- Sie haben Datensätze archiviert oder für den öffentlichen Zugang veröffentlicht
- mit Datenschutzbestimmungen und Compliance-Anforderungen vertraut sind
- Techniken zum Auffinden und Zugreifen auf relevante Datensätze eingesetzt haben
- in Ihren Projekten Datenmanagement-Tools oder -Software eingesetzt haben
- Sie haben andere in effektiven Datenverwaltungspraktiken geschult oder angeleitet.

Data Management und ein Data Management Plan (DMP)

Was sind Daten?

- Daten sind unkörperliche **Fakten, Zeichen** und **Symbole**.

Was sind Daten?

- Daten sind unkörperliche **Fakten, Zeichen** und **Symbole**.
- Wir definieren sie oft nach ihrer **Quelle** (z.B. *Verwaltung, Geschichte, Medizin usw.*) und ihren **Formaten** (z.B. *Zahlen, Text, Standbilder, Geodaten, Audio, Video und Software.*)

Was sind Daten?

- Daten sind unkörperliche **Fakten, Zeichen** und **Symbole**.
- Wir definieren sie oft nach ihrer **Quelle** (z.B. Verwaltung, Geschichte, Medizin usw.) und ihren **Formaten** (z.B. Zahlen, Text, Standbilder, Geodaten, Audio, Video und Software.)
- Sie wird als Grundlage der **Wissenshierarchie** unter der DIKW-Pyramide betrachtet.



Daten in der politischen Analyse

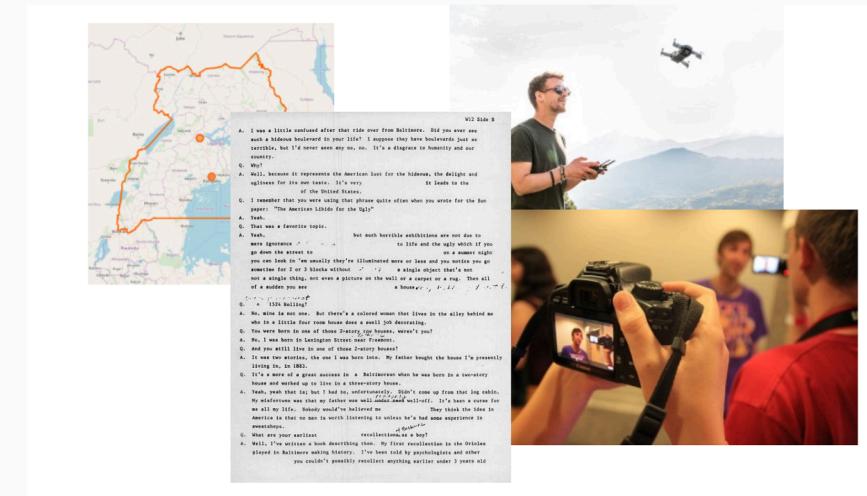
Ähnlich wie in der akademischen Forschung können Sie sich als **Politikanalysten auf eine breite Palette von Materialien** stützen, von strukturierten numerischen Datensätzen bis hin zu Interviews, Feldnotizen und Dokumenten, die für ethnografische Feldstudien gesammelt wurden,

Quantitative

startdate	startdate2	startdate3	lastupdate	lastupdate2	lastupdate3	collectorcreator
13599588513	26-Sep-2013 15:28:33	2013-09-26 15:29:01.440	13599589090	26-Sep-2013 15:38:10	2013-09-26 15:37:45.728	DIGITA08
13598878173	18-Sep-2013 10:09:33	2013-09-18 10:08:51.200	13598878822	18-Sep-2013 10:20:22	2013-09-18 10:19:46.560	DIGITA07
13598879269	18-Sep-2013 10:27:49	2013-09-18 10:28:30.848	13598879836	18-Sep-2013 10:37:16	2013-09-18 10:37:15.136	DIGITA07
13598879940	18-Sep-2013 10:39:00	2013-09-18 10:39:26.208	13598880525	18-Sep-2013 10:48:45	2013-09-18 10:48:10.496	DIGITA07
	18-Sep-	2013-09-18		18-Sep-	2013-09-18	

Viewing rows 6 through 9 of 1511

Qualitative



Unabhängig von der "Art" werden wir es in der Politik wahrscheinlich weitgehend mit der **Erhebung von Daten über Einzelpersonen** zu tun haben (d. h. i-pink[Forschung am Menschen]).

Alle Daten, die es Ihnen ermöglichen, eine Person zu **identifizieren**, werden als personenbezogene Daten eingestuft. In der Allgemeinen Datenschutzverordnung (DSGVO) sind personenbezogene Daten alle Informationen, die sich auf eine **identifizierte** oder **identifizierbare** natürliche Person, die als „betroffene Person“ bezeichnet wird, beziehen.¹.

Sensible personenbezogene Daten

Bestimmte personenbezogene Daten können besonders schützenswert sein, wenn sie Informationen preisgeben, die erhebliche *Risiken* für die *Grundrechte* und *Freiheiten* der betroffenen Person darstellen können. Im Zusammenhang mit der DSGVO können dies sein:

- Ethnische Herkunft;
- Politische Meinungen;
- Religiöse oder philosophische Überzeugungen;
- Gewerkschaftzugehörigkeit;
- Genetische Daten;
- Biometrische Daten;
- Daten über die Gesundheit;
- Daten über das Sexualleben oder die sexuelle Identität einer Person.

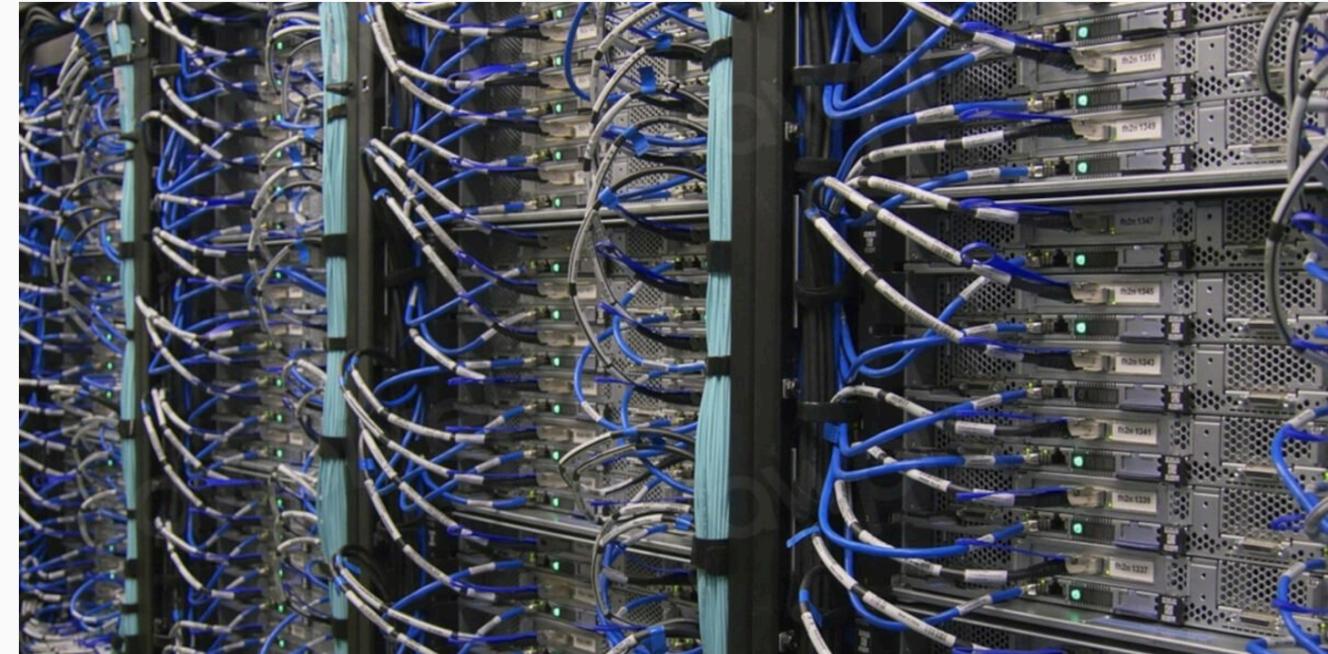


¹Die Datenschutz-Grundverordnung (DSGVO) gilt nur für die Daten von lebenden Personen. Daten, die nicht als personenbezogene Daten gelten, fallen nicht unter die Datenschutzvorschriften, obwohl es dennoch ethische Gründe für den Schutz dieser Informationen geben kann.

Research Data Management ist wie
“*health care*” für Ihre Daten:

- sie vor Schaden bewahrt,
- macht sie nutzbar und
auffindbar.

Dazu gehören Strategien, Prozesse
und Maßnahmen zur Erhaltung der
Datenqualität, der
Interpretierbarkeit von
Forschungsergebnissen und der
(Wieder-)Verwendbarkeit Ihrer



Der *Bedarf RDM* sorgsam zu überdenken steigt mit *sensiblen personenbezogenen Daten* an.

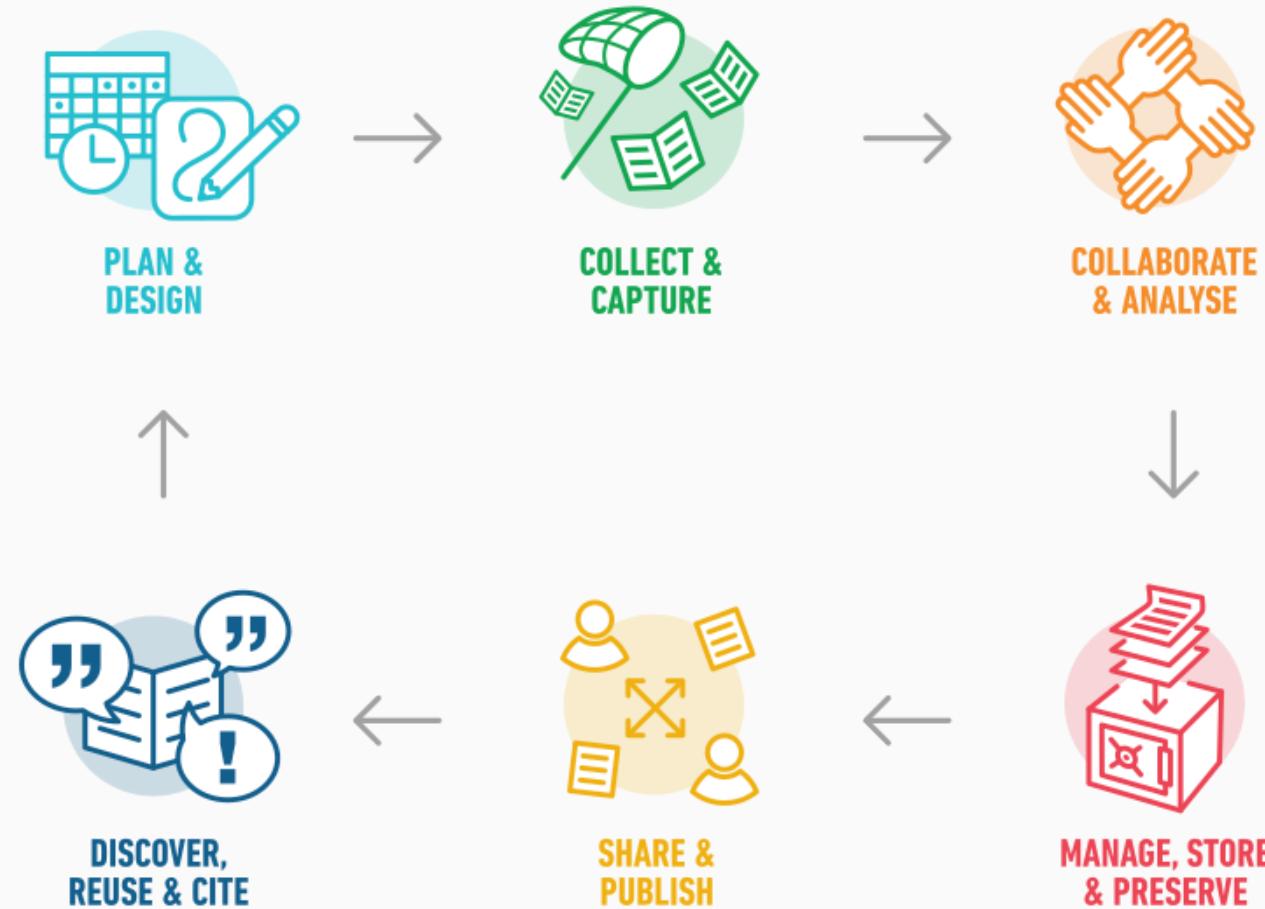
Research Data Management ist wie
“*health care*” für Ihre Daten:

- sie vor Schaden bewahrt,
- macht sie nutzbar und
auffindbar.

Dazu gehören Strategien, Prozesse
und Maßnahmen zur Erhaltung der
Datenqualität, der
Interpretierbarkeit von
Forschungsergebnissen und der
(Wieder-)Verwendbarkeit Ihrer



Research Data Management (RDM) Lifecycle



Datenmanagementpläne sind ein wichtiges Instrument zur Strukturierung des Datenmanagements Ihres Projekts.

Definiert Strategien, Maßnahmen und Verantwortlichkeiten für:

- die Verarbeitung und Validierung,
- Speicherung und Schutz,
- Bewahrung und gemeinsame Nutzung

Ihre Daten während des gesamten Datenzyklus.

Immer mehr *Finanzierungspartner verlangen diesen.*

Bestandteile eines Datenmanagementplans

1. Zusammenfassung der Daten
2. FAIR Daten
 - Auffindbar
 - Zugänglich
 - Interoperabel
 - Wiederverwendbar
3. Zuteilung von Ressourcen
4. Sicherheit der Daten
5. Ethische Aspekte
6. Sonstiges*

DMP component	Issues to be addressed
1. Data summary	<ul style="list-style-type: none">- State the purpose of the data collection/generation- Explain the relation to the objectives of the project- Specify the types and formats of data generated/collected- Specify if existing data is being re-used (if any)- Specify the origin of the data- State the expected size of the data (if known)- Outline the data utility: to whom will it be useful
FAIR data: 2.1. Making data findable, including provisions for metadata	<ul style="list-style-type: none">- Outline the discoverability of data (metadata provision)- Outline the identifiability of data and refer to standard identification mechanism. Do you make use of persistent and unique identifiers such as Digital Object Identifiers?- Outline naming conventions used- Outline the approach towards search keywords- Outline the approach for clear versioning- Specify standards for metadata creation (if any). If there are no standards in your discipline describe what type of metadata will be created and how

Data management plan (DMP) (cont.)

DMP component	Issues to be addressed
FAIR data: 2.2 Making data openly accessible	<ul style="list-style-type: none">- Specify which data will be made openly available? If some data is kept closed provide rationale for doing so- Specify how the data will be made available- Specify what methods or software tools are needed to access the data? Is documentation about the software needed to access the data included? Is it possible to include the relevant software (e.g. in open source code)?- Specify where the data and associated metadata, documentation and code are deposited- Specify how access will be provided in case there are any restrictions
FAIR data: 2.3. Making data interoperable	<ul style="list-style-type: none">- Assess the interoperability of your data. Specify what data and metadata vocabularies, standards or methodologies you will follow to facilitate interoperability.- Specify whether you will be using standard vocabulary for all data types present in your data set, to allow inter-disciplinary interoperability? If not, will you provide mapping to more commonly used ontologies?

DMP Komponente	Zu behandelnde Fragen
FAIR data: 2.4. Verbesserung der Wiederverwendung von Daten (durch Klärung der Lizzenzen)	<ul style="list-style-type: none">- Legen Sie fest, wie die Daten lizenziert werden, um die weitestmögliche Wiederverwendung zu ermöglichen- Legen Sie fest, wann die Daten zur Wiederverwendung verfügbar gemacht werden. Falls zutreffend, geben Sie an, warum und für welchen Zeitraum ein Datenembargo erforderlich ist- Legen Sie fest, ob die im Projekt erzeugten und/oder verwendeten Daten von Dritten nutzbar sind, insbesondere nach Ende des Projekts? Wenn die Wiederverwendung einiger Daten eingeschränkt ist, erklären Sie warum- Beschreiben Sie die Prozesse zur Qualitätssicherung der Daten- Legen Sie die Dauer fest, für die die Daten wiederverwendbar bleiben
3. Zuweisung von Ressourcen	<ul style="list-style-type: none">- Schätzen Sie die Kosten für die Umsetzung Ihrer Daten nach FAIR-Prinzipien. Beschreiben Sie, wie Sie diese Kosten decken wollen- Identifizieren Sie klar die Verantwortlichkeiten für das Datenmanagement in Ihrem Projekt- Beschreiben Sie die Kosten und den potenziellen Wert der langfristigen Aufbewahrung
4. Datensicherheit	<ul style="list-style-type: none">- Behandeln Sie Datenwiederherstellung sowie sichere Lagerung und Übertragung sensibler Daten
5. Ethische Aspekte	<ul style="list-style-type: none">- Im Kontext der Ethikprüfung, des ethischen Abschnitts des DoA und der ethischen Lieferungen zu behandeln. Schließen Sie Verweise und verwandte technische Aspekte ein, falls diese nicht von den vorgenannten abgedeckt werden
6. Sonstiges	<ul style="list-style-type: none">- Beziehen Sie sich auf andere nationale/finanzierende/branchenspezifische/abteilungsspezifische

Beschützen

Forschungsethik

Die moralischen **Prinzipien** und **Handlungen**, die Forschung leiten und gestalten

In der Politikanalyse ist eng mit der Forschungsethik in den Sozialwissenschaften verbunden.

- Ursprünglich ein Modell des 'Patientenschutzes' aus der medizinischen Forschung.
- Heute hat es einen breiteren Anwendungsbereich, einschließlich:
 - Berücksichtigung von *Nutzen, Risiken und Schäden*
 - *für alle Personen*, die mit der Forschung in Verbindung stehen und davon betroffen sind
- Weist **Verantwortlichkeiten** den Forschern und Analysten zu (z.B. *rechtlicher Rahmen*).



Nürnberg Code (1947)

Ethische Richtlinien für die Vorbereitung und Durchführung von medizinischen, psychologischen und anderen Experimenten am Menschen:

"Die freiwillige Zustimmung der Versuchsperson ist absolut unerlässlich [...], ohne dass ein Element der Gewalt, des Betrugs, der Täuschung, der Nötigung, der Übervorteilung oder einer anderen Form von Zwang oder Nötigung vorliegt [...], sollte über ausreichende Kenntnisse und ein ausreichendes Verständnis der Elemente des betreffenden Gegenstands verfügen, um sie in die Lage zu versetzen, eine **einsichtige und aufgeklärte Entscheidung** zu treffen."



Ethikkommission (IRB) und ethische Gremien

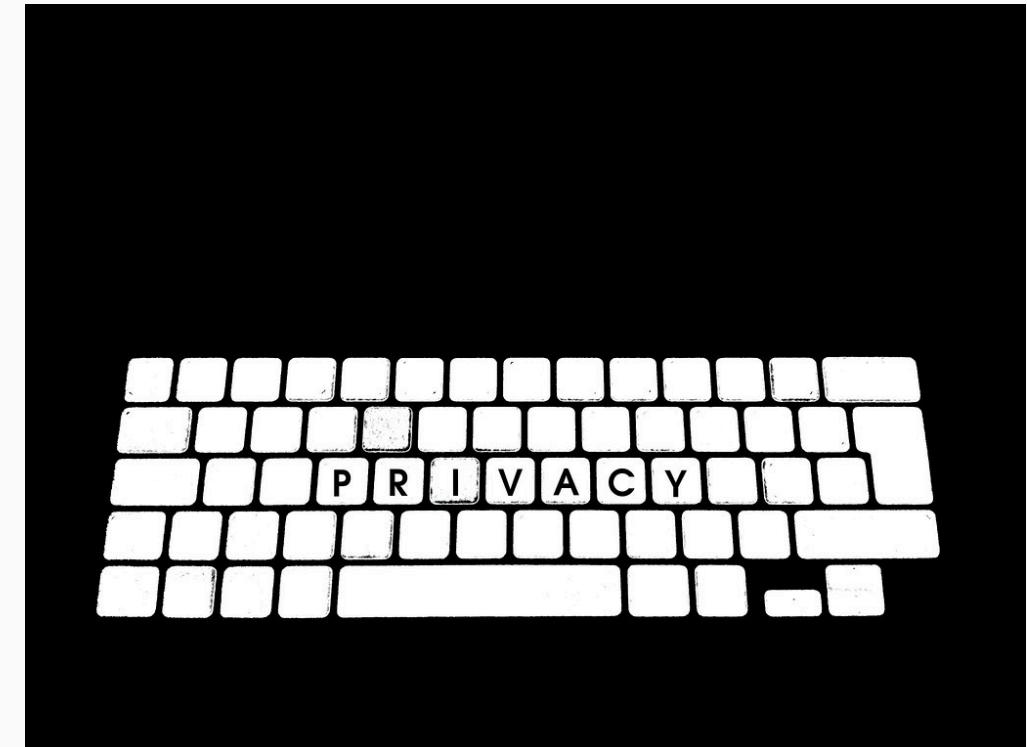
- Verwaltungsorgane, die eingerichtet wurden, um die **Rechte und das Wohlergehen** der Forschungsteilnehmer zu schützen
- Oft innerhalb der akademischen Einrichtung angesiedelt
- Überprüft Vorschläge und empfiehlt Änderungen und Verbesserungen
- *Kann unethische Projekte durch Verweigerung der Genehmigung stoppen*



Was ist "Datenschutz"?

Datenschutz

- Teil des Grundrechts *Recht auf Privatsphäre* (oder 'informationelle Freiheit').
- In der Forschung kann es zu Spannungen zwischen den Grundrechten kommen:
 - Freiheit der Forschung vs. Freiheit der persönlichen Information.
- „Privatsphäre ist ein persönlicher Lebenszustand, der durch Abgeschiedenheit von der Öffentlichkeit und damit durch Abwesenheit von deren Kenntnisnahme gekennzeichnet ist“ (Neethling 2005).
- **Kern:**
 - Verhinderung der unerwünschten Offenlegung von persönlichen Informationen oder des Missbrauchs solcher Informationen.



Hinter Daten befinden sich Individuen...



Article GDPR	Topic	Meaning
Art 5 (1) (a)	Lawfulness	Data must be processed in a legal way (Art 6) and transparent for ‘data subjects’; no surprises or covert activities.
	Fairness	
	Transparency	
Art 5 (1) (b)	Purpose limitation	Data may only be collected “for specified, explicit and legitimate purposes and not further processed in a manner that is incompatible with those purposes”; research exemption: research seen as in line with initial purposes.
Art 5 (1) (c)	Data minimisation	Limit amount of data collected.

Art GDPR	Topic	Meaning
Art 5 (1) (d)	Accuracy	Data collected for a given purpose should be kept correct and deleted or corrected without delay if necessary.
Art 5 (1) (e)	Storage limitation	Research exemption: longer period, if “appropriate technical and organizational measures” are implemented.
Art 5 (1) (f)	Integrity	Protected against “unauthorized or unlawful processing and against accidental loss, destruction or damage”.
	Confidentiality	
Art 5 (2)	Accountability	Controller (or processor) in charge and liable.

General Data Protection Regulation

- In Kraft seit 25. Mai 2018 (*fast auf den Tag genau sechs Jahre 😎*)
 - 99 Artikel und 173 Erwägungsgründe
 - Gilt unmittelbar
 - Soll das Datenschutzrecht EU-weit harmonisieren
 - aber, etwa 150 „Öffnungsklauseln“ oder Ausnahmen...
- GDPR (faktisch) integriert in eine Normenhierarchie



¹ Schauen wir uns diesen Artikel an <https://www.bbc.com/news/uk-wales-politics-58395974>

Organisieren, Dokumentieren, Verarbeiten und Speichern

Article 6 (1) GDPR	Examples
a) Consent	
b) Performance of a contract	Employment contracts; "two-sided obligational relationship" (e.g., sales contract); membership in association; <i>contract with the person concerned!</i>
c) Compliance with a legal obligation	Legal obligation to process data; e.g. documentation obligations under commercial law or notification obligations under social security law.

Article 6 I GDPR	Examples
d) Protection of vital interests	<i>Immediate threat</i> to vital interests of people exists; e.g., humanitarian emergencies and catastrophes.
e) Public interest or exercise of official authority	Tasks in the public interest; e.g., health, or justice and law enforcement.
f) Legitimate interests	Most important legal basis; advantage > flexibility; higher risks of data processing, because sole base on the initiative of the controller; e.g., data in companies, internet, credit agencies.

Art. 4 Abs. 11 GDPR

""Einwilligung' der betroffenen Person ist **jede Willensbekundung, die ohne Zwang, für den konkreten Fall, in Kenntnis der Sachlage und unmissverständlich erfolgt** und mit der die betroffene Person durch eine Erklärung oder eine eindeutige bestätigende Handlung zu verstehen gibt, dass sie **mit der Verarbeitung der sie betreffenden personenbezogenen Daten** einverstanden ist;“

Bedingungen für die Zustimmung

- Schriftform nicht mehr erforderlich.
- Bedingungen für die Einwilligung (Artikel 7 GDPR):
 - Nachweispflicht (Abs. 1)
 - Trennungsgebot (Abs. 2)
 - Leichte Widerrufbarkeit zu jeder Zeit (Abs. 3)
 - Erhöhte Anforderung an die Freiwilligkeit (Abs. 4)
- Verschärfung für Minderjährige unter 14 Jahren (Art. 8 GDPR).
- Wichtig: Erhebung „besonderer Kategorien“ personenbezogener Daten gemäß Artikel 9 DSGVO „verboten“, es sei denn, es gibt eine Rechtsgrundlage.
- **Zentral:** *Einwilligung muss dokumentiert werden!*

Können wir Daten ohne Einwilligung verarbeiten? (in der Forschung)

Die School

Dürfen **Forscher** Daten verarbeiten, die ohne Einwilligung gesammelt wurden?

- *Es kommt darauf an...*
- Es könnte notwendig sein, nachträglich eine Einwilligung einzuholen, nachdem die Daten gesammelt wurden.
- Dieser Schritt kann übersprungen werden, wenn "die Bereitstellung solcher Informationen unmöglich ist oder einen unverhältnismäßigen Aufwand bedeuten würde" (Art 14 Abs 5 Lit b).
- Aber das ist keine Standardlösung!
- **Unverhältnismäßig bedeutet wirklich unverhältnismäßig.**



Einwilligung nachträglich ist eine Interessenabwägung

„Interessenabwägung“ - Checkliste:

- Legitimes Interesse des Forschers?
- Ist die Datenverarbeitung notwendig?
 - Gibt es mildere Mittel, um das Forschungsziel zu erreichen?
- Kann die betroffene Person der Verarbeitung widersprechen?
- Sind diese Daten mit anderen Daten verknüpft (oder verknüpfbar)?
- Wie lange werden die Daten gespeichert?
- Wie viele Personen werden auf die Daten zugreifen?
- Gehören die Forschungsteilnehmer zu einer schutzbedürftigen Gruppe?
- Müssen die Forschungssubjekte mit der Verarbeitung ihrer Daten rechnen?

Überlegen Sie genau: Wenn die Interessen der Versuchspersonen schwerer wiegen als Ihre, ist die Datenverarbeitung unzulässig

Technical and Organizational Measures (TOMs)

- **Die Sicherheit unserer Daten**

TOMs sollten so konzipiert sein, dass „Datenschutzgrundsätze wie die Datenminimierung wirksam umgesetzt und die erforderlichen Garantien in die Verarbeitung integriert werden, um die Anforderungen dieser Verordnung zu erfüllen und die Rechte der betroffenen Personen zu schützen“. (Art 25 Par 1 GDPR).



Measure	Implementation
Data avoidance	Only collect as much data as necessary.
Data separation	Keep e.g. survey data (audio files, transcripts, ...) and contact data separate.
Admission control to building / office	E.g. lock office after leaving.
Access control to computer	E.g. protect PC with personal password.
Access control to files	E.g. agree on access to the electronic filing system and keep it as low as possible.

Pseudonymisierung

(Art. 4 Abs. 1 Nr. 5 GDPR):

„Verarbeitung personenbezogener Daten in einer Weise, dass die personenbezogenen Daten ohne Hinzuziehung zusätzlicher Informationen nicht mehr einer spezifischen betroffenen Person zugeordnet werden können, sofern diese zusätzlichen Informationen gesondert aufbewahrt werden und technischen und organisatorischen Maßnahmen unterliegen, die gewährleisten, dass die personenbezogenen Daten nicht einer identifizierten oder identifizierbaren natürlichen Person zugewiesen werden;“

Anonymisierung

(DSGVO Erwägungsgrund 26):

„Informationen, die sich nicht auf eine identifizierte oder identifizierbare natürliche Person beziehen oder auf personenbezogene Daten, die so anonymisiert wurden, dass die betroffene Person nicht oder nicht mehr identifizierbar ist.“

Mit Anonymisierung gibt es keine Möglichkeit, die wahre Identität festzustellen

Bei der Pseudonymisierung bleibt die Möglichkeit, die wahre Identität festzustellen

- Das ist eine Frage, auf die es **keine klare Antwort** gibt.
- Betrachten Sie den **Risiko-basierten Ansatz** der GDPR.
- Technische und organisatorische Maßnahmen (TOMs) helfen Ihnen, das Risiko in Schach zu halten.
- Wenn Sie mit großen Datensätzen arbeiten, sollten Sie Mittel zur **Statistischen Offenlegungskontrolle** (z. B. k-Anonymität) in Betracht ziehen.
- Wenn Sie unsicher sind, wenden Sie sich an einen **Datenmanagement- und Rechtsexperten!!!**



Kennt jemand den Unterschied zwischen *Datenschutz* und *Datensicherheit*?

Datensicherheit ist umfassender und nicht unbedingt personenbezogen.

- Verschlüsselung
- Starke Passwörter
- Zugriffsrechte
- Backups



Verschlüsselung erhält die Sicherheit von Daten

- Verwendet einen Algorithmus zur Umwandlung von Informationen
- Benötigt einen „Schlüssel“ zum Entschlüsseln

Mit Verschlüsselung können

- Daten zu übertragen
- Speichern von Daten (Backups)
- Auf entfernten Datenträgern

Tools. z.B. 7Zip, Gpg4win, Veracrypt



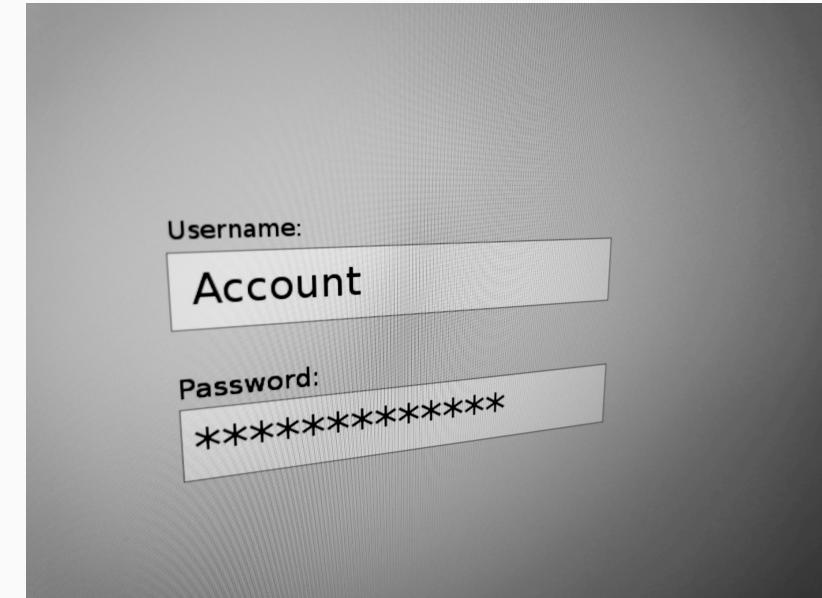
Hier finden Sie einige zusätzliche Informationen zur Verschlüsselung vom [Bundesamt für Sicherheit in der Informationstechnik](#)

Ein starkes Passwort hat:

- acht bis fünfzehn Zeichen oder sogar mehr
- eine zufällige Verteilung der Zeichen

Kombinieren Sie

- Großbuchstaben: A - Z
- Kleinbuchstaben: a - z
- Ziffern: 0-9
- Sonderzeichen: !"#\$%&'()*+,-./; etc.



Verwenden Sie einen 'Pass-Satz' anstelle eines Passworts!

Hier sind einige zusätzliche Informationen über Passwort-Manager vom Bundesamt für Sicherheit in der Informationstechnik 

Risiken:

- Technische Defekte
- Katastrophen
- Diebstahl
- Vergesslichkeit

Strategien

- Speicherung auf sicheren Servern mit automatischer regelmäßiger Sicherung
- Sicherung wichtiger Dateien in mindestens drei Kopien auf räumlich getrennten Datenträgern

Einrichtung der Datensicherung (3-2-1-Regel)

- Mindestens **3** Kopien einer Datei
- Auf mindestens **2** verschiedenen Medien
- Mindestens **1** davon ist remote

Testen Sie die Datenwiederherstellung zu Beginn und in regelmäßigen Abständen

Schützen Sie Ihre (sensiblen) Daten:

- Hardware (z.B. separater abschließbarer Raum)
- Dateiverschlüsselung
- Passwortsicherheit
- Mindestens zwei Personen sollten Zugang zu Ihren Daten haben

Hier finden Sie zusätzliche Informationen zur Datensicherung vom [Bundesamt für Sicherheit in der Informationstechnik](#) 

Schauen wir uns ein paar Fälle an

- **Was ist passiert?**
- **Warum war das ein Problem?**
- **Was hätte geschehen sollen?**

UK Information Commissioner's Office (ICO)

Fragen?
