

# Vereinbarkeit von Forschungsprozess & Datenmanagement

## Forschungsdatenmanagement nüchtern betrachtet



Vortrag im Rahmen des 107. Bibliothekartags  
TK 7: lehren & unterstützen: Forschungsprozess und Datenmanagement  
Berlin, 13. Juni 2018

# Ein traditionelles Rezept wird neu interpretiert!



Foto: Wikimedia Commons CC-SA 4.0



Foto: Colourbox: 2515474

## 1. Erstellung eines Forschungskonzeptes

- Entwicklung der Fragestellung
- Auswahl der Untersuchungsmaterialien und Methoden

## 2. Entwicklung des Arbeitsprogramms

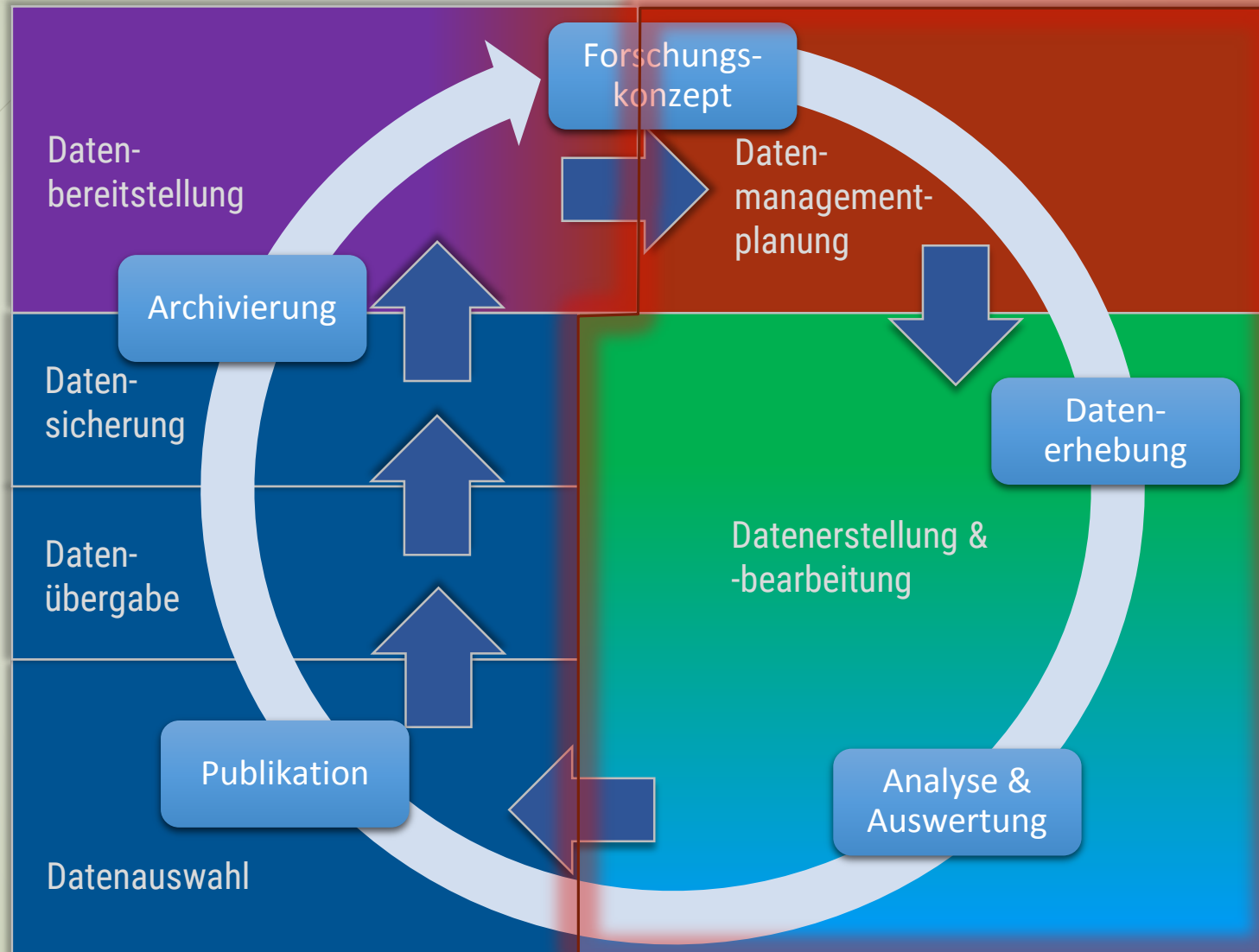
- Definition der Arbeitsschritte für die Beantwortung der Forschungsfrage und Erreichung der Ziele
- Unter Berücksichtigung des Datenmanagements



# Forschungsprozess



# Datenmanagement



# Von Untersuchungsmaterialien zu Forschungsdaten

Beispiele aus der Praxis



# Ziele und Arbeitsprogramm

## 1. Fragen und Ziele

### ➤ Fragen

- Wer steht wie intensiv mit der Person in Kontakt und in welcher Beziehung stehen sie zueinander und wie entwickelt sie sich?
- Gibt es bei der Verwendung von Wortbildungsarten Unterschiede zwischen mündlicher und schriftlicher Kommunikation?

### ➤ Ziele

- Online Publikation



**Digitale Projekte müssen stärker vom Ende her geplant werden!**

# Ziele und Arbeitsprogramm

## 2. Materialien und Methoden

- Untersuchungsmaterialien
  - Korrespondenzen von und an die Person (Briefe, Postkarten, Telegramme) (zentraler Quellenkorpus)
  - Archivalisches Schriftgut (Zeitungsartikel, Protokolle, Berichte) (ergänzend)
  - Kommentartexte im Internet, Gesprächstranskripte, Emails, Zeitungsartikel
- Methoden
  - Qualitative Auswertung (Wer spricht über welche Themen?)
  - Quantitative Auswertung (Korrespondenzkontakte, Themenschwerpunkte, Häufigkeit von bestimmten Wortbildungsarten)

**Quellen und Methoden sind die Grundlage für die Entwicklung des Datenmodells**



# Ziele und Arbeitsprogramm

## 3. Arbeitsschritte

### 1. Recherche und Erfassung

- Metadaten der Briefe zum Teil in Kalliope vorhanden
- Gesprächstranskripte vom IDS Mannheim
- Kommentartexte von diversen Plattformen per Web-Crawler aggregiert
- Restliche Quellen/Texte durch Archiv- und Bibliotheksrecherchen erschliessen und digitalisieren

### 2. Erschließung und Aufbereitung

- Tiefergehende Beschreibungen des jeweiligen Materials (z.B. Beziehungen untereinander)
- Volltexte generieren (Transkriptions-, Editionsrichtlinien)



# Ziele und Arbeitsprogramm

## 3. Arbeitsschritte

### 3. Analyse und Auswertung

- Inhaltsanalyse
- Auszeichnung von Personen, Orten
- Erfassung von Personenbeziehungen
- Häufigkeitsabfragen

### 4. Ergebnispublikation

- Online-Publikation

**Entwicklung des Datenmodells und  
der Workflows gehen Hand in Hand**





# Workflow- und Datenmodellentwicklung

## 1. Recherche und Erfassung

- Welche Informationen (Metadaten) zu den Materialien werden benötigt und in welcher Form?
  - Zitierbarkeit
  - Auffindbarkeit
  - Methodisch bedingte Metadaten
- Berücksichtigung von Standards notwendig ? (z.B. EAD, CEI)



# Beispiel Erfassungsmaske in FuD



## "Allgemein" - Brief DEM1-br-0001: Max Mustermann

Absender (Transkript)	Willy Brandt	<input type="checkbox"/>
Absender	Brandt, Willy	<input type="checkbox"/>
Empfänger (Transkript)	Präsident	<input type="checkbox"/>
Empfänger	Kennedy, John F.	<input type="checkbox"/>
Dokumenttitel	Brief des Regierenden Bürgermeisters von Berlin, Willy Brandt, an den amerikanischen Präsidenten John F. Kennedy, 16. August 1961	<input type="checkbox"/>
Datum (JJJJ-MM-TT)	1961-08-16	<input type="checkbox"/>
Datum	16. August 1961	<input type="checkbox"/>
Datum (Verifikation)	gesichert	<input type="checkbox"/>
Ausstellungsort	Deutschland→Berlin	<input type="checkbox"/>
Ausstellungsort (Verifikation)	gesichert	<input type="checkbox"/>
Empfängerort	Vereinigte Staaten von Amerika→Washington D.C.	<input type="checkbox"/>
Empfängerort (Verifikation)	gesichert	<input type="checkbox"/>
Seitenumfang		<input type="checkbox"/>

# Workflow- und Datenmodellentwicklung

## 1. Recherche und Erfassung

- Wie können die Daten erfasst werden?
  - manuelle oder (teils-)automatisierte Erhebung der Metadaten (Import)
- Welche Hilfsmittel stehen in den Bibliotheken und Archiven zur Verfügung?
  - Elektronische Findbücher, Kataloge (Import)
  - Internetverbindung



# Workflow- und Datenmodellentwicklung

## 2. Erschließung und Aufbereitung

- Digitalisierung der Materialien(z.B. photomechanisch durch externen Service)
- Volltextgenerierung
  - OCR-Software möglich
  - Manuelle Transkription (ggf. softwareunterstützt)

## 3. Analyse und Auswertung







- Forschungsfragen- bzw. methodenspezifische Aspekte, z.B.
  - Qualität von Beziehungen zwischen Untersuchungsmaterialien
  - Eindeutige Identifizierung von Personen und Orten
    - Aufnahme zusätzlicher Informationen zu Person (Beziehungen, Geburts-, Sterbedatum, Berufe etc.) und Orten (Namensvarianten, Geo-Koordinaten etc.)



# Beispiel Erfassungsmaske Person (Allgemein) in FuD

FuD Demo Version 2 - Max Mustermann















Inventarisierung Export Dokumentverwaltung Recherche Analyse Freigabe Einstellungen Werkzeuge Hilfe

**INVENTARISIERUNG**

Allgemein Namen Ereignisse Beziehungen Zusammenführung Literatur Kommentar Text Anhang Arbeitsschritte







**"Allgemein" - Person DEM1-ps-0001: Max Mustermann**

Erfassungssprache	Deutsch	 	<input type="checkbox"/>
<b>Kategorisierung</b>			
Projektkategorie	Politiker	  	<input type="checkbox"/>
FuD Kategorie	Person	 	<input type="checkbox"/>
Geschlecht	männlich	 	<input type="checkbox"/>
URN Normdatensatz	<a href="http://d-nb.info/gnd/119187752">http://d-nb.info/gnd/119187752</a>		<input type="checkbox"/>
Personenindex	Kennedy, John F.	 	<input type="checkbox"/>
<b>Name</b>			
Art des Namens	Geburtsname	 	<input type="checkbox"/>
(Vor)Name(n)	John Fitzgerald		<input type="checkbox"/>
Nachname	Kennedy		<input type="checkbox"/>
Namenszusatz			<input type="checkbox"/>

# Beispiel Erfassungsmaske Person (Ereignisse) in FuD

FuD Demo Version 2 - Max Mustermann

Inventarisierung Export Dokumentverwaltung Recherche Analyse Freigabe Einstellungen Werkzeuge Hilfe






**INVENTARISIERUNG**

Allgemein **Namen** Ereignisse Beziehungen Zusammenführung Literatur Kommentar Text Anhang Arbeitsschritte

**"Allgemein" - Person DEM1-ps-0001: Max Mustermann**

Nr. of Events 5

Ereignisse

Lebensdaten · Geburt · 1917-05-29 · JJJJ-MM-TT · Vereinigte Staaten von Amerika→Massachusetts→Brookline · Kennedy, Joseph Patrick; Kennedy, Rose Fitzgerald	 
Lebensdaten · Tod · Am 22. November 1963 wurde John F. Kennedy auf einer Wahlkampfreise gegen 12:30 Uhr an der Dealey Plaza, einem Platz in Dallas, Texas mit mehreren Gewehrschüssen während einer Fahrt im offenen Wagen durch die Innenstadt von Dallas ermordet. · 22.11.1963 gegen 12:30 · Text · 1963-11-22 · gesichert · JJJJ-MM-TT · Vereinigte Staaten von Amerika · Texas · Dallas · Dealey Plaza · Oswald Lee	 
Politische Laufbahn · Wahl zum Präsidenten der Vereinigten Staaten · 1961-01-20 · JJJJ-MM-TT · Vereinigte Staaten von Amerika→Washington D.C.	 
Politik · Treffen mit Willy Brandt · 1961-03-13 · gesichert · JJJJ-MM-TT · Vereinigte Staaten von Amerika→Washington D.C.→Weißes Haus · Brandt, Willy · Berlin-Krise	 

# Workflow- und Datenmodellentwicklung

## 3. Analyse und Auswertung

- ▶ Inhaltliche Erschließungen → Operationalisierung des methodischen Vorgehens
  - ▶ Analysekategorien und ggf. deren gegenseitige Abhängigkeit und Eigenschaften festlegen
  - ▶ Vorgehen zur Entwicklung des Vokabulars definieren
  - ▶ Definition von Eigenschaften der Beobachtungen
  - ▶ Grenzen der Operationalisierung identifizieren
- ▶ Auswertung
  - ▶ Bilden die verwendeten Tools das methodische Vorgehen 1:1 ab?
  - ▶ Wie müssen die Abfragen formuliert werden?



# Beispiel Analyse in FuD (methodenspezifische Konfiguration)

FuD Demo Version 2 - Max Mustermann

Inventarisierung Export Dokumentverwaltung Recherche Analyse Freigabe Einstellungen Werkzeuge Hilfe

Registerrückbau Anzeigeanalyseobjekte  
Index nach Analysefreigaben nach Dokumenten nach Ir  
 mit Analyseobjekten  aufgeklappt  eigene  fremde  alles  alle Dokumente  Mappe  Aktuelles Dokument  nur

## KATEGORIEN - INDEX

Freie Analyse

Mauerbau Personenindex Raumindex Sachindex Begegnungen

aktiv  Suche

- Gipfelkonferenz (1) DEM1-br-0001
- Archiv
- Berlin-Krise
- Blockade (1) DEM1-br-0001
- Defensive (1) DEM1-br-0001
- Dreimächtestatus (1) DEM1-br-0001
- Gebäude
- Kapelle
- Mauerbau
- Militär (1) DEM1-br-0001
- Ostberlin (1) DEM1-br-0001
- Ostblock (2) DEM1-br-0001
- Reichsbürger
- Souveränität (1) DEM1-br-0001
- Vier-Mächte-Status (1) DEM1-br-0001
- Volksentscheid

## DOKUMENT [DEM1-BR-0001] ALLGEMEIN

[DEM1-br-0001] Allgemein: Brief (Max Mustermann)

Absender: Brandt, Willy  
Empfänger: Kennedy, John F.  
Dokumenttitel: Brief des Regierenden Bürgermeisters von Berlin, Willy Brandt, an den amerikanischen Präsidenten John F. Kennedy, 16. August 1961  
Datum: 1961-08-16 (Anzahl Wörter: 666)

Indiziert von Max Mustermann, fud@uni-trier.de

Sehr verehrter Herr Präsident!

Nach den Entwicklungen der letzten drei Tage in meiner Stadt möchte ich Ihnen in diesem persönlichen und informellen Schreiben einige der Gedanken und Gesichtspunkte mitteilen, die mich bewegen.

Die Maßnahmen des Ulbricht-Regimes, gestützt durch die Sowjetunion und den übrigen Ostblock, haben die Reste des Vier-Mächte-Status nahezu völlig zerstört. Während früher die Kommandanten der alliierten Mächte in Berlin bereits gegen Paraden der sogenannten Volkarmee protestierten, haben sie sich jetzt mit einem verspäteten und nicht sehr kraftvollen Schritt nach der militärischen Besetzung des Ostsektors durch die Volksarmee begnügen müssen. Die illegale Souveränität der Ostberliner Regierung ist durch Hinnahme anerkannt worden, soweit es sich um die Beschränkung der Übergangsstellen und des Zutritts zum Ostsektor handelt. Ich halte dies für einen ernsten Finschnitt in der Nachkriegsgeschichte dieser



# Beispiel Analyse in FuD (methodenspezifische Konfiguration)

FuD Demo Version 2 - Max Mustermann

Inventarisierung Export Dokumentverwaltung Recherche Analyse Freigabe Einstellungen Werkzeuge Hilfe

Registeraufbau  
Index  
 mit Analyseobjekten  aufgeklappt

Anzeige Analyseobjekte  
nach Analysefreigaben  
 eigene  fremde  alles

nach Dokumenten  
 alle Dokumente  Mappe  Aktuelles Dokument

nach Ir  nur

Indexeintrag ändern

**Index Sachindex**

**Übergeordnete Indexeinträge**

Indexeintrag

Format

GND

Synonym

Merkmale

FuD-Dokumentlink

Verbindungstyp

Weblink

Bemerkung

DEM1-BR-0001] ALLGEMEIN

Allgemein: Brief (Max Mustermann)

Willy Brandt, John F. Kennedy, Präsident des Regierenden Bürgermeisters von Berlin, Willy Brandt, an den Präsidenten John F. Kennedy, 16. August 1961 (Anzahl Wörter: 666)

Mustermann, fud@uni-trier.de

Präsident!

lungen der letzten drei Tage in meiner Stadt möchte ich Ihnen in meinem informellen Schreiben einige der Gedanken und Meinungen mitteilen, die mich bewegen.

des Ulbricht-Regimes, gestützt durch die Sowjetunion und den Rest der Vier-Mächte-Status nahezu völlig zerstört.

Die Kommandanten der alliierten Mächte in Berlin bereits gegen die Volksarmee protestierten, haben sie sich jetzt mit einem sehr kraftvollen Schritt nach der militärischen Besetzung des Ostsektors begnügen müssen. Die illegale Souveränität der DDR ist durch Hinnahe anerkannt worden, soweit es sich um die Übergangsstellen und des Zutritts zum Ostsektor handelt.

ihnen einen ersten Einschnitt in der Nachkriegsgeschichte dieser

# Beispiel Analyse in FuD (methodenspezifische Konfiguration)

FuD Demo Version 2 - Max Mustermann

Inventarisierung Export Dokumentverwaltung Recherche Analyse Freigabe Einstellungen Werkzeuge Hilfe

Registerrückbau Anzeige Analyseobjekte  
Index nach Analysefreigaben nach Dokumenten nach Ir  
 mit Analyseobjekten  aufgeklappt  eigene  fremde  alles  alle Dokumente  Mappe  Aktuelles Dokument  nur

## KATEGORIEN - INDEX

Freie Analyse

Mauerbau **Personenindex** Raumindex Sachindex

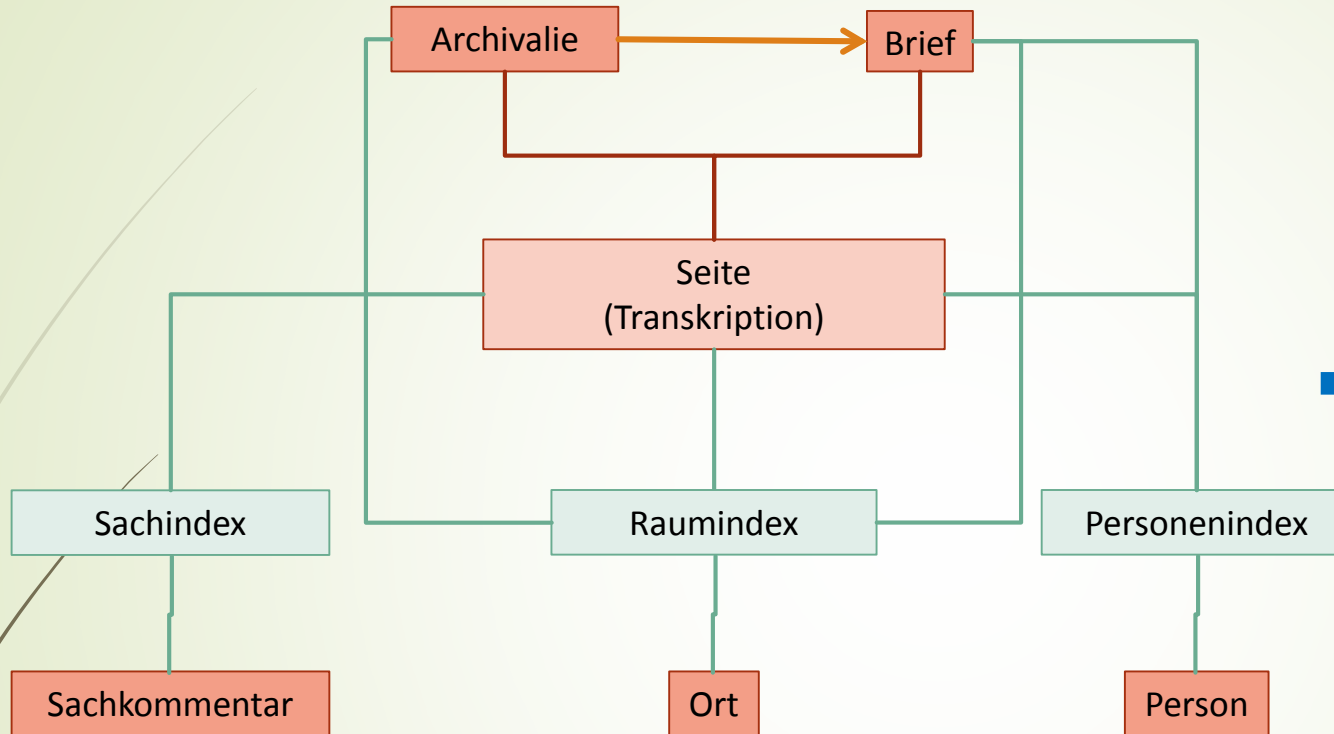
aktiv

- Gipfelkonferenz (1) DEM1-br-0001
- Archiv
- Berlin-Krise
- Blockade (1) DEM1-br-0001
- Defensive (1) DEM1-br-0001
- Dreimächtestatus (1) DEM1-br-0001
- Gebäude
- Kapelle
- Mauerbau
- Militär (1) DEM1-br-0001
- Ostberlin (1) DEM1-br-0001
- Ostblock (2) DEM1-br-0001
- Reichsbürger
- Souveränität (1) DEM1-br-0001
- Vier-Mächte-Status (1) DEM1-br-0001
- Volksentscheid

## Analyseobjekt bearbeiten

Wortart	Analogiebildung	
Lexikalisierung	teeinologiker	
Grundform	Substantiv	<input type="button" value="↑↓"/> <input type="button" value="≡+"/> <input type="button" value="≡"/>
Struktur	Okkasionell	<input type="button" value="↑↓"/> <input type="button" value="≡+"/> <input type="button" value="≡"/>
Morpheme	binär	<input type="button" value="↑↓"/> <input type="button" value="≡+"/> <input type="button" value="≡"/>
fkt.-sem. Klasse	UK Analogiebildung → Alkoholiker --> Teeinologiker	<input type="button" value="↑↓"/> <input type="button" value="≡+"/> <input type="button" value="≡"/>
WB-Bed.		<input type="button" value="≡+"/> <input type="button" value="≡"/>
Stil		<input type="button" value="≡+"/> <input type="button" value="≡"/>
Fremdspr.		<input type="button" value="≡+"/> <input type="button" value="≡"/>
Anm.	bei Struktur und Morphemen nicht sicher	<input type="button" value="≡+"/> <input type="button" value="≡"/>

# Beispiel eines Datenmodells



## Legende



## ► DOKUMENTATION!!!!

► Wie wird ein Objekt repräsentiert und in welchem logischen Zusammenhang steht es zu den anderen Objekten im Datenmodell?

► Jedes Objekt im Modell braucht

► eine Beschreibung (Metadaten),

► i.d.R. eine inhaltliche, logische Beziehung zu mindestens einem anderen Objekt

und

► einen Workflow (der beschreibt und festlegt, wo die Informationen herkommen und wie sie erfasst werden)

# Beobachtungen & Reflexionen



# Geisteswissenschaftler\*innen bringen das Handwerkszeug zum Forschungsdatenmanagement mit!

- ▶ Kritischer Umgang mit Untersuchungsmaterialien befähigt zur bewussten und v.a. kritischen Auseinandersetzung mit den Materialien (Quellenkritik)
  - ▶ Befähigt zur Erstellung von Metadatenschemata, da man seine Materialien detailgenau beschreibt und klassifiziert
- ▶ Fähigkeit komplexe, gesellschaftliche Systeme multiperspektivisch zu rekonstruieren und dennoch werden trotz fehlender Überlieferungen Lücken logisch argumentativ geschlossen
  - ▶ Befähigt zur Entwicklung von detaillierten Workflows zur Datengenerierung, wobei relativ kleinteilig jeder Arbeitsschritt logisch auf einem vorhergehenden aufbauen muss



# Aber warum tun sich die Geisteswissenschaftler\*innen dann so schwer damit?

- ▶ Wegen der Technik, die sie meinen nicht verstehen bzw. nicht beherrschen zu können
  - ▶ Dafür gibt es andere Experten, die die definierten Prozesse den Maschinen beibringen
- ▶ Es kostet viel Zeit und Mühe, die eingeübten Arbeits- und Vorgehensweisen auf die digitalen Bedingungen umzustellen

## Ja, aber ...

- ▶ wir ziehen daraus auch Vorteile für unsere zukünftigen Forschungen: Nicht nur durch die vereinfachte und ressourceneffizientere Nachnutzbarkeit, Auffindbarkeit und Zusammenführung von Datenkorpora, sondern es können sich sowohl die bereits existierenden als auch noch entstehenden Technologien zu Nutzen gemacht und damit das Methodenrepertoire erweitert werden.
- ▶ den Digital Turn kann man nicht mehr aufhalten, jetzt sind die Wissenschaftler\*innen nicht nur aufgerufen, sondern haben die Chance den Turn aktiv mitzugestalten.



# Aber warum tun sich die Geisteswissenschaftler\*innen dann so schwer damit?

- ▶ Der zusätzliche Aufwand für die Bereitstellung nachnutzbarer Forschungsdaten wird nicht honoriert.

## Ja, aber ...

- ▶ für eine neue Wissenschaftskultur, die die Bereitstellung qualitätsgesicherter Forschungsdaten wertschätzt, sind die Wissenschaftler\*innen in erster Linie selbst verantwortlich.



Vielen Dank für Ihre Aufmerksamkeit !



### **Literaturhinweis**

Gisela Minn und Marina Lemaire: [Forschungsdatenmanagement in den Geisteswissenschaften](#). Eine Planungshilfe für die Erarbeitung eines digitalen Forschungskonzepts und die Erstellung eines Datenmanagementplans. Trier Juli 2017. <[urn:nbn:de:hbz:385-10715](#)>



## Kontakt

Marina Lemaire

E-Mail: [marina.lemaire@uni-trier.de](mailto:marina.lemaire@uni-trier.de)

[www.esciences.uni-trier.de](http://www.esciences.uni-trier.de)

[www.facebook.com/esciencesunitrier](https://www.facebook.com/esciencesunitrier)

[www.twitter.com/esciencetrier](https://www.twitter.com/esciencetrier)

