

Von Papyri bis Pixels

Praxisbeispiele für Forschungsdatenmanagement
in den Geisteswissenschaften

Vortrag im Rahmen des
4. virtuellen Forschungsdatentags
der Rhein-Main Universitäten

Marina Lemaire

Universität Trier. Servicezentrum eSciences



KI-generiertes Bild via Midjourney

Im Namen des 4Memory-Konsortiums



- Erstellung eines Forschungskonzeptes
 - Erkenntnisinteresse und Fragestellung
 - Auswahl der Untersuchungsmaterialien
 - Identifikation der Untersuchungsmethoden
- Entwicklung des Arbeitsprogramms
 - Definition der Arbeitsschritte für die Beantwortung der Forschungsfrage und Erreichung der Ziele



Unter Berücksichtigung
des Datenmanagements

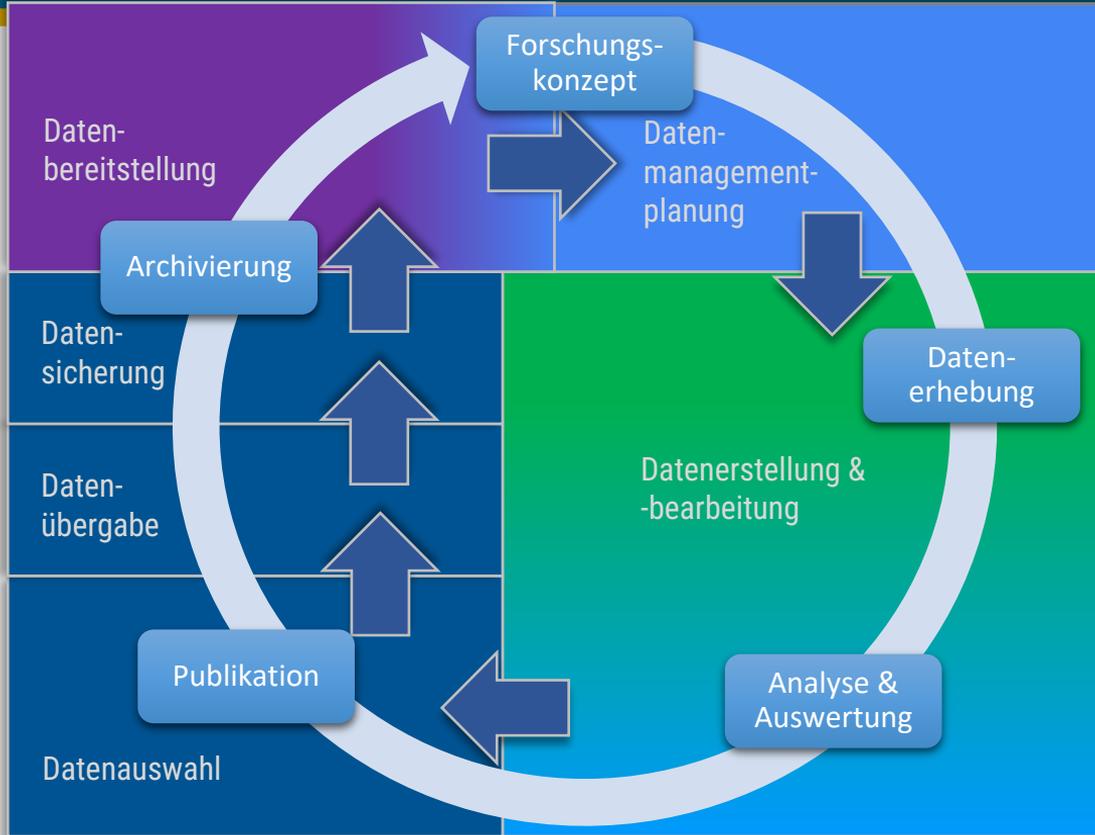
- Quellen liegen digital vor
- Nutzung digitaler Werkzeuge in allen Schritten des Arbeits- und Forschungsprozesses
- Erzeugung von digitalen Forschungsdaten

Aber: Kein Wandel des grundsätzlichen Erkenntnisinteresses der geisteswissenschaftlichen Forschung

Wer, wie, was – wieso, weshalb, warum?

KI-generiertes Bild via Midjourney





MIDA DAS MODERNE INDIEN IN DEUTSCHEN ARCHIVEN 1706-1989
Von der DFG gefördertes Langfristprogramm

Suche zurücksetzen

Sortierung: Anzahl

Verzeichnungstufe

- Archivalie (1.399)
- Archivbestand (2.688)
- Archiv (80)

Archivsparte

- Staatliche Archive (43)
- Archive der Hochschulen sowie wissenschaftlicher Institutionen (13)
- Wirtschaftsarchive (8)

Merkliste

Studentenakten I: 1829-1913 [Stadtstudenten]	
Signatur	UAT, 40
Datum	01.01.1829 - 31.12.1913
Studentenakten I: 1829-1913 [Konviktooren]	
Signatur	UAT, 41
Datum	01.01.1829 - 31.12.1913
Studentenakten I: 1829-1913 [Seminaristen]	

Praxisbeispiel Datenveröffentlichung

Das moderne Indien in deutschen Archiven, 1706–1989 (MIDA)

- Bestände deutscher Archive zur Geschichte des modernen Indiens und zur deutsch-indischen Verflechtungsgeschichte systematisch erfassen und verschlagworten
- Bereitstellung der Datenbank für internationale Forschung
 - Anschlussfähigkeit an Archivdatenbanken (Standards)
 - Präsentationsplattform orientiert an Archivportal-D
- Pflege eines sukzessiv erweiterbares digitalen Archivführers



FDM muss stärker vom Ende her gedacht werden!

- Einheitliche Datenerfassung über diverse Forschungsprojekte hinweg, die die unterschiedlichen deutschen Archive für die Quellenrecherche aufsuchen (nicht in allen Internetanschluss möglich)
- Untersuchungsmaterialien = Archivalien
 - Detaillierte Erschließung auf Archiv, Bestands und ggf. Dokumentenebene



Quellen, Methoden und Rahmenbedingungen sind die Grundlage für die FDM-Strategieentwicklung

Recherche und Erfassung

- Manuelle Erfassung der Metadaten per einheitlicher Excelsheets
- Erfassung der notwendigen Metadaten orientiert am EAD und an den Bedarfen für das eigene Datenmanagement und die spätere Suche im Portal
- 3 Dokumenttypen: Archiv, Archivbestand und Archival und Festlegung des Datenmodells

Erschließung und Aufbereitung

- Erfassung per vorbereiteten Excelsheets, die die Datenmodelle abbilden durch Forschende in Einzelprojekten
- Regelmäßige Importe in die gemeinsame Datenbank durch Datenkurator
 - Harmonisierung der Daten
 - Anreicherung von Personendaten mit GND
 - Überprüfung der Daten durch einen dritten Projektmitarbeiter zur Qualitätssicherung



Entwicklung des Datenmodells und des Workflows gehen Hand in Hand

Publikation

- Präzisierung des Präsentationskonzeptes
 - Zielgruppenspezifische Suchstrategien
 - Darstellung der Dokumentinformationen
 - Navigation
 - Festlegung von Exportformaten
- Implementierung der Präsentationsplattform durch Serviceanbieter
 - Testen der Präsentationsumgebung und iterative Weiterentwicklung der Plattform



Ob alles problemlos umsetzbar ist, ist abhängig vom Datenmodell

Publikation

- Prüfung und Freigabe der Dokumente für die Publikation
- Release der Plattform (Rolling Publication)

- Inhalte für die Informationsseiten zur Plattform, z.B. ein kleines Manual, Nutzungsbedingungen etc. erstellen
- Redaktionsworkflows für den Archivführer festlegen

- Entwicklung eines Nachhaltigkeitskonzepts

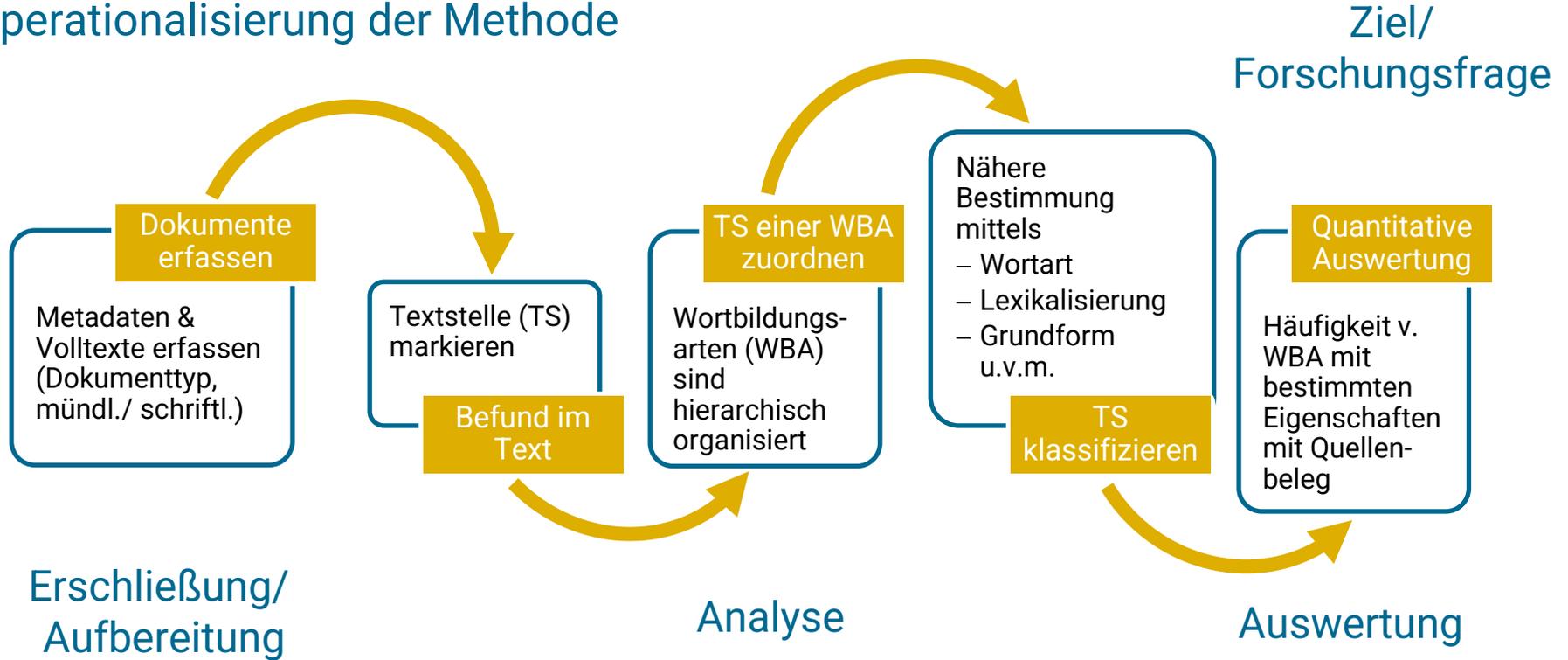
Archivierung

- Repository der SUB Göttingen nach Projektabschluss
- Alle Daten werden in den Formaten (XML(TEI), JSON, PDF, LaTeX) exportiert zzgl. DB-Dump und der Dokumentation des Datenmodells

The screenshot shows the 'formemory' software interface. On the left, a tree view lists 'Wortbildungsarten' (Word formation types) under 'Unmittelbare Konstante (UK)'. The tree includes categories like 'aktiv', 'Analogiebildung', 'Explizite Derivation', 'Komposition', 'Konversion', and 'Kurzwortbildung'. On the right, a table shows word counts for different types, with 'Analogiebildung' highlighted in green. A pop-up window titled 'Analyseobjekt bearbeiten' (Edit analysis object) is open, showing details for 'teeinologiker'. It includes fields for 'Wortart' (Substantiv), 'Lexikalisierung' (Okkasionell), 'Grundform' (Teeinologiker), 'Struktur' (binär), and 'Morpheme' (2). A note at the bottom of the pop-up reads 'Anm. bei Struktur und Morphemen nicht sicher' (Note: not sure about structure and morphemes). The word 'schmatzt' is visible on the right side of the pop-up window.

Praxisbeispiel für Operationalisierung der Analyse
**Wortbildung in Mündlichkeit und Schriftlichkeit. Beziehungen
 zwischen Wort- und Textbildung im geschriebenen und
 gesprochenen Deutsch**

Operationalisierung der Methode



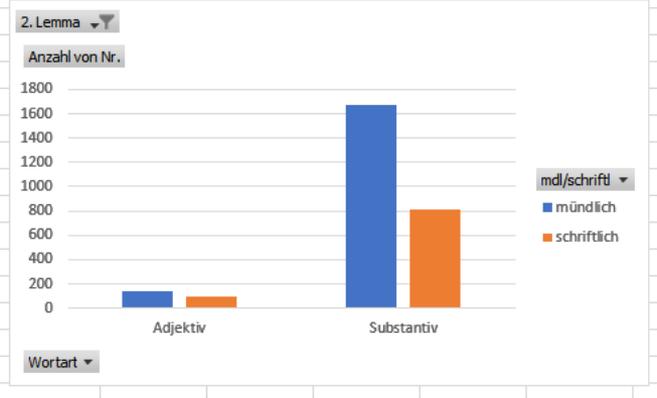
Umsetzung in FuD-Software

The screenshot displays the 'Freie Analyse' (Free Analysis) window in the FuD software. On the left, a tree view under 'Wortbildungsarten' (Word Formation Types) is expanded to 'Determinativkompositum (2226)'. A yellow callout box labeled 'TS einer WBA zuordnen' (Assign a TS to a WBA) points to this category. The main window shows a search result for 'Cappuccinopulver' with its morphological analysis: 'Determinativkompositum cappuccinopulver', 'Wortart Substantiv', 'Lexikalisierung Okkasionell', 'Grundform Cappuccinopulver', 'Struktur binär', 'Morpheme 2', 'UK Determinativkompositum → Substantiv + Substantiv', 'fkt.-sem. Klasse', 'WB-Bed.', 'Stil', and 'Fremdspr. italienisch'. A yellow callout box labeled 'Befund im Text' (Finding in Text) points to the word in the search text: '... das_s cappuccinopulver mit viel milch (,) das wird'. A third yellow callout box labeled 'TS klassifizieren' (Classify TS) points to the 'Determinativkompositum' classification. The bottom of the window has buttons for 'Übernehmen' (Take over), 'Bereich ändern' (Change area), and 'Abbrechen' (Cancel).

Export aus FuD und Auswertung in Excel

Nr.	Ident.	Inhaber	Analyseobj.	1. Lemm	2. Lemm	Wortart	lexikalisiert	Grundform	Struktur	Morphen	UK	Anm.	Titel	Medium	
0	WB-td-0001	Stumpf, Sören	cappuccinop	Komposition	Determinativ	Substantiv	Okkasionell	Cappuccinop	binär		2	DK&N&Substantiv + Subs	Alltagsgespr	Gespräch zw	
1	WB-td-0001	Stumpf, Sören	türklinke	Komposition	Determinativ	Substantiv	Usuell	Türklinke	binär		2	DK&N&Substantiv + Subs	Alltagsgespr	Gespräch zw	
2	WB-td-0001	Stumpf, Sören	johannis (.)	Komposition	Determinativ	Substantiv	Usuell	Johannisbee	binär		2	DK&N&Substantiv + Subs	Alltagsgespr	Gespräch zw	
3	WB-td-0001	Stumpf, Sören	walnuss	Komposition	Determinativ	Substantiv	Usuell	Walnuss	binär		2	DK&N&Substantiv + Subs	Alltagsgespr	Gespräch zw	
4	WB-td-0001	Stumpf, Sören	wal[nuss	Komposition	Determinativ	Substantiv	Usuell	Walnuss	binär		2	DK&N&Substantiv + Subs	Alltagsgespr	Gespräch zw	
5	WB-td-0001	Stumpf, Sören	waln]uss	Komposition	Determinativ	Substantiv	Usuell	Walnuss	binär		2	DK&N&Substantiv + Subs	Alltagsgespr	Gespräch zw	
6	WB-td-0001	Stumpf, Sören	walnuss	Komposition	Determinativ	Substantiv	Usuell	Walnuss	binär		2	DK&N&Substantiv + Subs	Alltagsgespr	Gespräch zw	
7	WB-td-0001	Stumpf, Sören	bioekese	Komposition	Determinativ	Substantiv	Okkasionell	Biokeks	binär		2	DK&N&Konf	bio als Konfi	Alltagsgespr	Gespräch zw
8	WB-td-0001	Stumpf, Sören	bioeks[e	Komposition	Determinativ	Substantiv	Okkasionell	Biokeks	binär		2	DK&N&Konf	bio als Konfi	Alltagsgespr	Gespräch zw
9	WB-												Alltagsgespr	Gespräch zw	
10	WB-												Alltagsgespr	Gespräch zw	

Anzahl von Nr.		Spaltenbeschriftungen		
Zeilenbeschriftungen		mündlich	schriftlich	Gesamtergebnis
Adjektiv		142	98	240
Substantiv		1669	813	2482
Gesamtergebnis		1811	911	2722



- Was repräsentiert ein Datenobjekt? (Zeitschriftenartikel, Interview, Person, ...)?
 - in welcher Beziehung 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 (Metadaten, Digitalisate, Aufzeichnungen etc.) herkommen, wie sie erfasst werden, in welchem Format, mit welchem System und wer dafür verantwortlich ist



KI-generiertes Bild via Midjourney

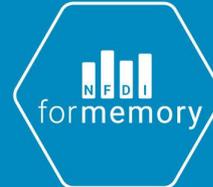
KI-generiertes Bild via Midjourney



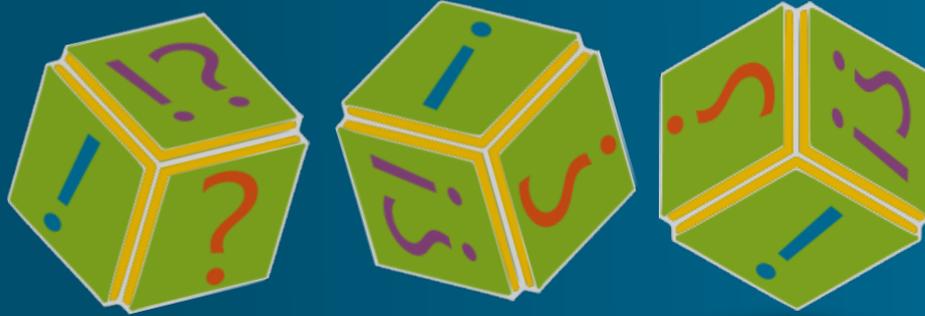
Von Büchern zu Bytes

Die digitale Transformation in der
Geschichtswissenschaft – Lernen,
Lehren und Forschen im Zeitalter
der Digitalisierung

Eine Vortragsreihe von:



<https://4memory.de/dataliteracy/>



Herzlichen Dank für Ihre Aufmerksamkeit