

# STRUKTUR UND KONTEXT: MODELLE DIGITALER FORSCHUNGSDATEN IN DEN GEISTESWISSENSCHAFTEN

Maxi Kindling (AGEP CMS / IBI)

Magisterarbeit

# **Möglichkeiten der Strukturmodellierung: Eine exemplarische Zusammenführung funktionaler Anforderungen an die Bereitstellung digitaler Forschungsdaten für ausgewählte geisteswissenschaftliche Disziplinen**

Arbeit zur Erlangung des akademischen Grades einer Magistra Artium

# Gliederung

- Themenfelder
- Ausschnitt aus der Masterarbeit: Hermeneutik und Besonderheiten in GW
- Modelle
- Bedeutungsfindung und Kontextualisierung
- Strukturmodellierung
- Ergebnisse
- Abschlussbetrachtung
- Fazit
- Anschluss Themen

# Thema: Forschungsdaten

- „junges“ Forschungsfeld  
(insbes. in Deutschland)
  - viele offene Fragen

# Themenfelder

- Definitionsprobleme hinsichtlich der Begriffe ‚Forschungsdaten‘ oder auch ‚Primärdaten‘
- Konzepte ‚eScience‘ bzw. ‚eResearch‘, Bereitstellung in Virtuellen Forschungsumgebungen
- Forschungsdaten in der wissenschaftlichen Wertschöpfung

# Themenfelder

- Einschränkung auf Geisteswissenschaften
  - Aber: Es gibt DIE Geisteswissenschaften nicht.
  - Daher: Beschränkung auf **Hermeneutik** als der Interpretation von Forschungsgegenständen

# Motivation zur Untersuchung

- Häufige Aussage im Forschungsdatendiskurs:  
„Mit der langfristigen Speicherung, Bereitstellung und Publikation von Forschungsdaten werden neue Forschungsansätze möglich. Es können neue Forschungsfragen gestellt werden.“

Was verbirgt sich dahinter (in der Theorie)?

# Annahmen für die Untersuchung

## Arbeitsdefinitionen:

*Gegenstände geisteswissenschaftlicher Forschung sind alle menschlichen 'Hervorbringungen' wie Sprache oder kulturelle Artefakte wie Musik und Kunst.*

*Als digitale Forschungsdaten für die Geisteswissenschaften werden [...] all jene genuin digitalen Daten sowie die digitalen Repräsentationen analoger Quellen bezeichnet, die gemäß der [...] Definition Gegenstände geisteswissenschaftlicher Forschung und damit Diskursinhalte sein können.*

## Abgrenzung zu Naturwissenschaften:

- Gegenstandsbereich der Forschung
- Form des Zugangs (Methode)
- Verhältnis von Instrument und Gegenstand



# Annahmen für die Untersuchung

Einschränkung auf die hermeneutische Methode

- Hermeneutik bedeutet die Auslegung von Zeichen = Interpretation / Bedeutungszuweisung
  - Achtung: Dies ist eine Arbeitsdefinition!
- verstehende „Quellenwissenschaft“ vs. erklärende Naturwissenschaft

# Annahmen für die Untersuchung

## Hermeneutik und Zeichen

- alles Symbolische hat Zeichencharakter und kann daher untersucht werden
  - sehr breite Auslegung
- ein Akteur interpretiert Zeichen
- mit der Interpretation durch Akteur erhält ein Zeichen Bedeutung
- Semantic Web: Formalisierung der Zeichenauslegung

# Arbeitshypothese

In den hermeneutisch geprägten Disziplinen der Geisteswissenschaften besteht ein genauer zu bestimmendes Verhältnis zwischen dem Zugang zum Forschungsgegenstand und dem Forschungsgegenstand selbst. Dies wird insbesondere in Hinblick auf die digital vernetzten Forschungsdaten deutlich.

# Verschränkung von Instrument und Gegenstand

- Besonderer Gebrauch von Zeichen: Interpretation erfolgt **diskursiv**, d.h. nicht datengeleitet, sondern selbstreflexiv
  - wesentlich: Verknüpfung und Einordnung in Kontext (historisch, sprachlich...)
- **Instrument** (Verfassen einer Sekundärschrift) **kann gleichzeitig Gegenstand sein** (Publikation)
  - sowohl Instrument (Prozess der Erkenntnisproduktion) als auch Gegenstand (Erkenntnis) haben Zeichenfunktion

# Verschränkung von Instrument und Gegenstand

- **Gegenseitige Bezugnahme** auch formal:
  - Publikation verweist durch einen **Zeiger** (Link) auf das **Gezeigte** wie einen Forschungsgegenstand und stellt damit eine Relation her
  - Fragen, die einen Forschungsgegenstand betreffen, können selbst zum Untersuchungsgegenstand werden
  - in NW werden sprachliche Aussagen (Zeiger) über Daten (Gezeigtes) getroffen; in den hermeneutisch geprägten GW werden diese Sprachaussagen selbst zum Gegenstand der Untersuchung

# Zwischenfazit

- Es gibt eine Vielzahl von Betrachtungsmöglichkeiten für digitale Forschungsdaten:
  - Wissenschaftliche Wertschöpfung
  - Digital Curation: Digitale Objekte
  - Verhältnis Daten – Information – Wissen (Semiotik, Informatik, Informationwissenschaft)
  - **disziplinär (Gegenstand und Methode)**

# Modelle

- zahlreiche Betrachtungsmöglichkeiten führen zu Problemen in der Begriffsbestimmung digitaler Forschungsdaten
- Lösung: Modelle als Bezugspunkt und Grundlage der Verständigung (kognitiv wie technisch), da sie auf gemeinsamen Vorstellungen bzw. Vereinbarungen über einen Gegenstand basieren
- Was ist eigentlich ein Objekt? Was ist seine Struktur? Wie lässt diese sich beschreiben?
  - Eine pragmatische Sicht würde sagen, dass ein digitales Objekt dann digitales Objekt ist, wenn es nach einem bestimmten Modell als ein solches beschrieben werden kann.
- Aber auch Modelle entstehen anhand des Wissensschatzes verschiedener Disziplinen ...

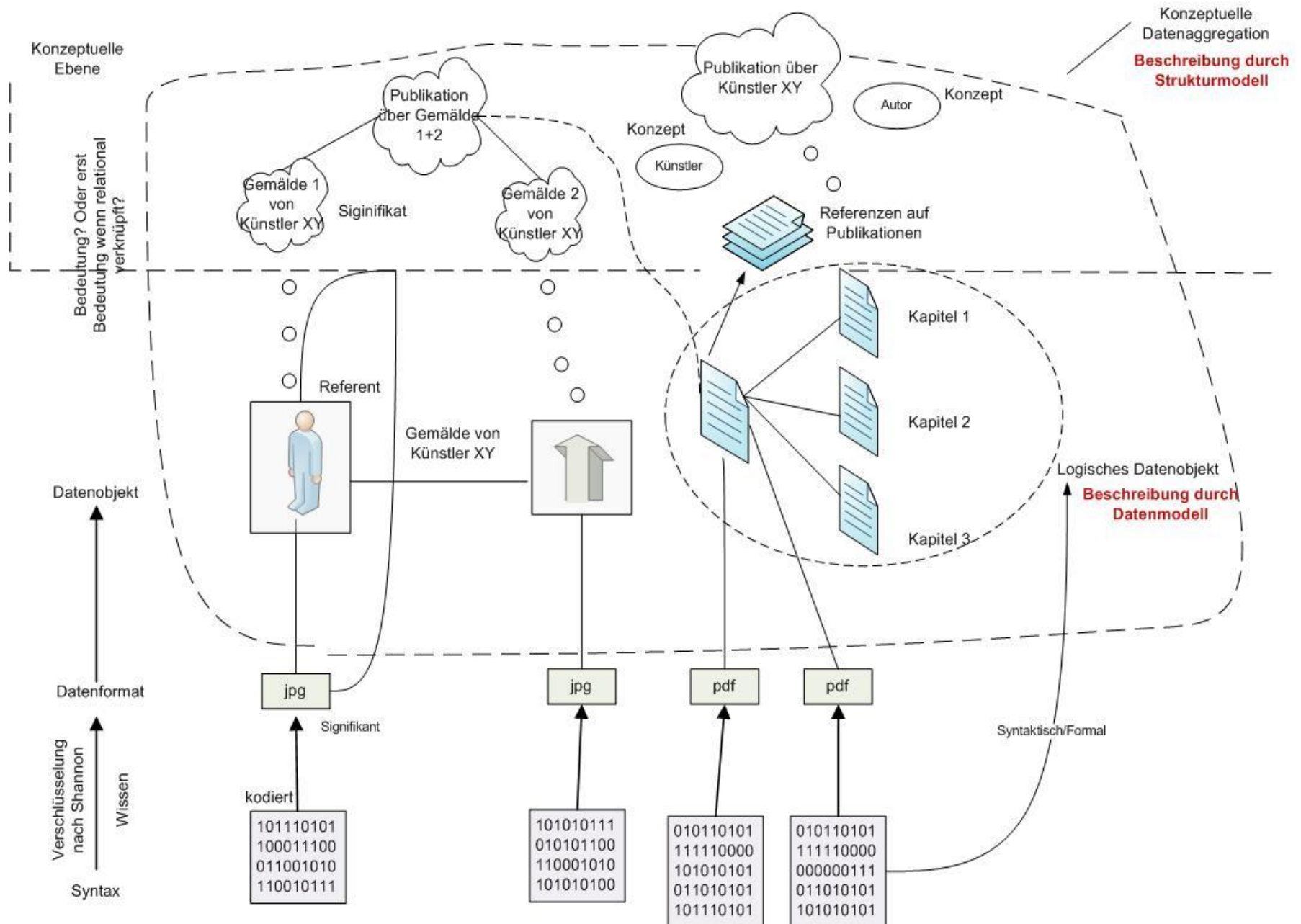
# Modelle

- Beispiel für die Abhängigkeit vom Betrachtungsmodell: OAIS-Modell
  - Digitale Informationsobjekte als Bedeutungsträger (semantisch wie sprachlich) können Wissen manifestieren (bspw. in Form einer Publikation), sind jedoch nicht selbst Wissen.
  - Es fehlt die pragmatische Komponente der Information: Interpretation im semiotischen Sinne durch Anwendung von Kontextwissen.
  - Wissen entsteht im Kontext. Die digitale Repräsentation per se entspricht der Qualität Daten und nicht „Information“ oder „Wissen“.



# Wissen im Kontext

- Forschungsprozess in den GW: Wissen entsteht durch die Verknüpfung von Informationsobjekten
- grundlegende Nutzungsszenarien der *humanities* nach John Unsworth:
  - Discovering
  - Annotating
  - Comparing
  - Referring
  - Sampling
  - Illustrating
  - Representing



# Bedeutungsfindung durch Kontextualisierung

- Bezugnahme auf Gegenstände und Instrumente durch die Herstellung von Verknüpfung
- Schwierigkeit: Verschränkung von Instrument und Gegenstand
- Kontext muss formaler (logischer) und konzeptueller (semantischer) modelliert werden

# Bedeutungsfindung durch Kontextualisierung

→ Lösung:

Modelle für komplexe digitale Objekte, die die Relationen zwischen Datenobjekten berücksichtigen (strukturell) und deren Relationen semantisch motiviert sein können

Voraussetzung: Persistente Identifizierung

# Strukturmodellierung

- Zusammenführung digitaler Objekte
- Darstellung der Verknüpfungen
- Nachnutzung der Datenaggregationen
- Nachvollziehbarkeit des Forschungsprozesses
- Transfer verteilter Datenbestände in digitalen vernetzten Umgebungen
- Praxis: Digitale Objekte in der Darstellung ihres Kontexts durch Daten- bzw. Objektmodelle wie OAI ORE für Webressourcen

# Ergebnisse

## These 1

Durch die gegenseitige Bezugnahme der Forschungsgegenstände sowie die Einbettung in neue inhaltliche Kontexte entsteht ein komplexes Beziehungsgefüge zwischen digitalen Objekten (auf der Ebene von Datenobjekten und aus konzeptueller Sicht).

# Ergebnisse

## These 2

Für die Darstellung dieser komplexen Beziehungsgefüge müssen die digitalen Objekte in ihrem Kontext modelliert werden.

# Ergebnisse

## These 3

Die strukturelle Verknüpfung digitaler Objekte schafft Wissen, da die Interpretation von Zeichen auf inhaltlich motivierten Relationen basiert, die damit Bedeutung tragen. Die Relationssemantik entsteht aus den Anforderungen der Wissenschaftsdisziplinen. Struktureller Kontext schafft demnach zugleich Bedeutungskontext.



# Ergebnisse

These 4 (nicht neu!)

Das nominalistische Zeichenverständnis der Informatik, in dem Verweise zwischen Begriffen und Objekten herstellt, muss erweitert werden: Objekte sind Bedeutungsträger und verweisen auf Objekte als Bedeutungsträger. Es entsteht ein komplexes Beziehungsgefüge, in dem der Begriff des Zeichens mehrdimensional betrachtet werden muss.

# Ergebnisse

These 5 (auch nicht neu!)

In der wissenschaftlichen Wertschöpfung im digitalen Raum wird aus dem vormals monolithischen Informationsobjekt ein generischer Bedeutungsträger (bspw. durch die Referenz von Bezugsquellen) . Das Dokument ist damit relativ, der Dokumentbegriff wird dekonstruiert.

→ Wir sprechen von Informationsaggregaten

# Abschlussbetrachtung

## Linked Data Web (Tim Berners-Lee)

- strukturierte Daten, die mithilfe von RDF formalisierten Verweise verknüpft werden
- die Verweise sind nicht lokal orientiert, sondern basieren auf der Identifizierung von digitalen Objekten
- Dies entspricht der Idee von digitalen Forschungsdaten in vernetzten Umgebungen.
- *„[...] data is relationships [...] the more things you have to connect, the more powerful is raw data [...]!“*
  - Bedeutung entsteht auf der Basis typisierter Relationen.
- Perspektive: Freie Daten im Web
- Alle Daten im Web können konzeptuell Forschungsdaten sein.



# Fazit

Ich habe **keine neuen Forschungsfragen** in den digitalen Geisteswissenschaft gefunden (und auch nicht gesucht), sondern untersucht, was ‚neue Forschungsfragen und –ansätze‘ **bedeuten kann** (Interpretation durch Heranziehen von Betrachtungsebenen und -modellen).

Konkrete neue Forschungsfragen sind nur durch die Wissenschaftler selbst zu entwickeln, die technischen Lösungen werden von der Informatik umgesetzt.

Das methodische Wissen der Geisteswissenschaften wird mit der praktischen Kompetenz der Informatik und den Erkenntnissen der **Bibliotheks- und Informationswissenschaft in der Funktion des Mediators** kombiniert.

# Anschlussthemen

## Merkmale wie Heterogenität und Komplexität

- heterogene Datentypen wie Video, Audio, Text, 3-D-Objekte, Bilder
- heterogene, z. T. nicht-standardisierte und proprietäre Datenformate
- heterogen in Bezug auf die Rezeptionswege (Lesegeräte etc.)
- komplex hinsichtlich der Struktur digitaler Sammlungen und Daten-Aggregation formal
- komplexe semantische Relationen zwischen den Daten
- hohe Zahl an Abstraktionsstufen: von Sprachdaten als empirisches Material (auch Textforschung, Text Mining) bis hin zur Untersuchung von Denkmodellen

# Anschlussthemen

## Weitere Aussagen zu digitalen FD in den GW

- meist eingeschränkter Zugriff auf Daten und dazugehörige Metadaten (oftmals auf physischen Datenträgern)
- durch mangelhafte und nicht interoperable Aufbereitung der Daten meist kaum nachgewiesen und auffindbar
- kaum Strategien für langfristige Verfügbarkeit
- kaum vorhandene oder nur unzureichende Werkzeuge zur Datenbearbeitung mit Ausnahme der Text- und Editionsforchung
- größer werdende Korpora an bspw. digitalem Sprachmaterial durch Digitalisierungsprojekte
- ...

# Anschlussthemen

## Theoretischer Natur

- Einbeziehung des Akteurs in Richtung „Pragmatic Web“
- Rolle von Data Sharing und Open Data im wissenschaftlichen Kommunikationsprozess
- rechtl. Fragen: Wem gehören Verknüpfungen?

oder zur Abwechslung mal praktisch:

- Anwendung von Strukturmodellierung
- Untersuchung der Semantik von Relationen
- Linked Data Web
- Beispiele wie ECHO, Hypernietzsche, Discovery...
- Untersuchung von wissenschaftlichem Diskurs



Vielen Dank!

Fragen jetzt oder auch gerne an:  
[maxi.kindling@ibi.hu-berlin.de](mailto:maxi.kindling@ibi.hu-berlin.de)