

HUMANITIES DATA CENTRE



Prämissen und Schlussfolgerungen bei der Angebotsentwicklung des Humanities Data Centres mit Hilfe des Konzepts des „Forschungsdatentyps“

Ulrike Wuttke

Stefan Buddenbohm, Claudia Engelhardt

16. September 2015



1. Einleitung

- Humanities Data Centre
 - Geisteswissenschaftliches Forschungsdatenzentrum
 - Designphase 2014-2016
 - Förderung durch das MWK
- Herausforderung
 - Kompromiss zwischen wissenschaftlichen Anforderungen und begrenzten Ressourcen
- Vortragsinhalt
 - Erfahrungen aus der Angebotsentwicklung

Gliederung

1. Einleitung
2. Charakteristika geisteswissenschaftlicher Forschungsdaten
3. Anforderungen aus Sicht der Wissenschaft
4. Anforderungen aus Sicht eines Infrastrukturanbieters
5. Abgeleitete Angebotsgestaltung über Forschungsdatentypen
6. Nutzungsdimensionen und initiales Angebotsportfolio des HDC
7. Schlussfolgerungen

2. Charakteristika geisteswissenschaftlicher Forschungsdaten

- nach wie vor große Bedeutung von Text
- zunehmend aber verbundene Objektsammlungen
 - d.h. nicht einfach nur Formatheterogenität
 - sondern: Verknüpfung verschiedenster Objekte und wissenschaftlicher Gehalt im Design (bspw. Anwendungsoberfläche)
 - i.d.R. keine dateigebundenen Daten wie bspw. in einem Publikationsrepository

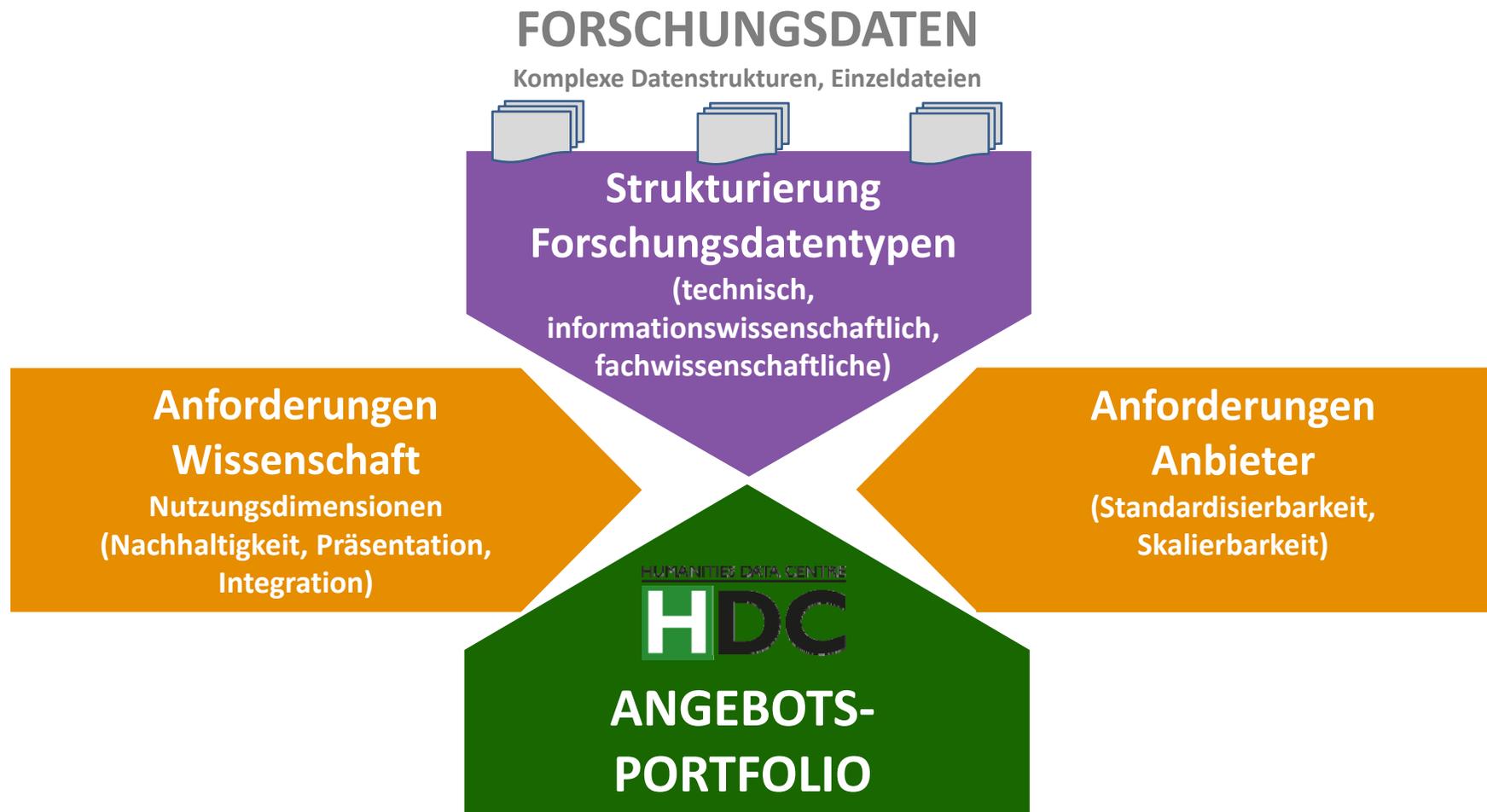
3. Anforderungen aus Sicht der Wissenschaft

- Verdichtung auf drei Nutzungsdimensionen:
 - Nachhaltigkeit
 - Präsentation
 - Integration

4. Anforderungen aus Sicht eines Infrastrukturanbieters

- Standardisierbarkeit
- Skalierbarkeit

5. Abgeleitete Angebotsgestaltung über Forschungsdatentypen



6. Nutzungsdimensionen und initiales Angebotsportfolio des HDC

- Anbindung an die vorgenannten wissenschaftlichen Nutzungsdimensionen
 - Nachhaltigkeit
 - Präsentation
 - Integration
- Zuordnung konkreter Angebote zu diesen Dimensionen (siehe folgende Abbildung)



Komplexe Datenstrukturen, Einzeldateien

BERATUNG

PORTAL

**GENERISCHE
ANWENDUNGS-
SYSTEME**

**ANWENDUNGS-
KONSERVIERUNG**

**DATENEXTRAKTION
STRUKTURBESCHREIBUNG**

REPOSITORY

LANGZEITARCHIVIERUNG

7. Schlussfolgerungen

- Angebot als Kompromiss zwischen wissenschaftlichen Anforderungen und zur Verfügung stehenden Ressourcen.
- Beschränkung als Grundlage für ein funktionsfähiges und finanzierbares Forschungsdatenzentrum.
- Nutzungsdimensionen: Nachhaltigkeit, Präsentation, Integration.

Vielen Dank für Ihre Aufmerksamkeit!



AutorInnen: Stefan Buddenbohm, Claudia Engelhardt, Ulrike Wuttke

E-Mail: kontakt@humanities-data-centre.org

Lizenz: 

Datum: 16. September 2015, Hamburg

Webseite: <http://humanities-data-centre.org/>

Präsentationsschicht

- Darstellung in Browser
- Daten können referenziert und in anderen Umgebungen eingebettet werden

Datenbank

- Zusammenführung von teils unvollständigen oder widersprüchlichen Informationen in eine (semi-)strukturierte Form

Quellen

- können unterschiedlichste Werke und Objekte sein (z.B. Bücher, Briefe, Plakate, Archiveinträge)
- typische Datensammlungen zu u.a.: Personen, Orten, Institutionen

Globale Suche: Go

Berliner Klassik Datenbanken | **Personen** | Bibliographie | Geselligkeit | Nationaltheater

Lebensdaten | Namensformen | Genealogie | Biographie | Werke / Literatur | Berlinaufenthalte | Register | Quellen

Burgsdorf, Friedrich August Ludwig von (23.03.1747 - 18.06.1802) Forstrat

Nachname	Burgsdorf	Vorname	Friedrich August Ludwig	Lexikal. Name
Titel		Adelstitel		Adelspraedikat von
Geburtsdatum	23.03.1747			Ungefähres Geburtsdatum
Geburtsort	Leipzig	Geburtsland		Staatsangehörigkeit
Geschlecht	m	Konfession		
Todesdatum	18.06.1802			Ungefähres

BK

- TestDBeintraege
- Theaterscans
- acl_users
- admin
- bibliographie
- bilder
- geselligkeit
- imghaupt
- personen
- theater

© Zope Corporation
Refresh

Contents View Properties Secu

BK to MySQL Project at /BK

Type	Name
<input type="checkbox"/>	Adelspraedikat_Administration.html (Adelspraedikat administrieren)
<input type="checkbox"/>	Adelstitel_Administration.html (Adelstitel administrieren)
<input type="checkbox"/>	BKPerson_anlegen.html
<input type="checkbox"/>	Benutzerdetails.html
<input type="checkbox"/>	Berlinaufenthalte_Liste.html (Liste der Personen mit all ihren Berlinaufenthalten)
<input type="checkbox"/>	Beruf_Administration.html (Berufe administrieren)
<input type="checkbox"/>	Each_Administration.html (Each administrieren)

Name	Haude- und Spenersche Zeitung
Datum	07.09.1805
Nummer	108
Seite	
Autor	ungez.
Text	Den 4ten September: Götz von Berlichingen mit der eisernen Hand, Schauspiel in fünf Akten von Göthe. Nach des Verfassers neuer Bearbeitung

Lang ist von dieser Umarbeitung die Rede gewesen; wir kennen sie nun endlich auch. Da das Stück so lang geworden ist, (es spielte bis 10 Uhr) und so mancherlei darin die Aufmerksamkeit fesselt, wird es schwer, gleich nach einmaliger Aufführung über sein Verhältniß zur ersten Edition bestimmt zu sprechen. Man wird natürlich auf die Beantwortung der Fragen gespannt seyn: Hat das Stück gewonnen? Von dem wodurch es vor dreißig Jahren so großes Glück gemacht hat, konnte ihm der Verfasser jetzt schwerlich etwas neues zulegen. Dazu gehörte erstlich das damals noch wenig gebrauchte Ritter-Kostüm; jetzt aber kömmt die Ritterzeit so häufig auf der Bühne vor, daß es Perioden gibt, wo die Hälfte aller Darstellungen in altdeutschen Kragen über die Bühne prunkt; ferner: eine Reihe verwegener, genialer Züge, welche zum Theil auch die (jetzt an Kotzebue viel getadelte) Natürlichkeit gewanter darstellen, als es jemals die

Publikationsschicht

- duale Darstellung von Editionsdigitalisat (bspw. JPEG) und Editionstranskription (HTML)
- Endnutzerseitig wird ein Browser vorausgesetzt



Transkription

- Erzeugung einer vernetzten Darstellung von Digitalisat und XML-basiertem Text sowie verschiedenen Annotationsebenen
- XML-Datenbank (eXist/SADE)



Quellen

- handschriftliche Arbeitsmanuskripte oder gedruckte Textdokumente
- je nach wissenschaftlicher Forschungsfrage werden Dokumente digitalisiert und erfasst



TRANSCRIPTION

P_7T

1 Einleitung

2 Naturwissenschaft (*Philosophia naturalis*) ist die Wissenschaft

3 von der Eigenschaften und den Bewegungsgesetzen der Materie (des

4 **Beweglichen** im Raum) überhaupt. - Principien *a priori* von dieser Wissen-

5 schaft sind entweder die mathematische oder die *metaphysischen*. (überwicht Beyde unter-

6 scheiden sich von der *mathematischen* Naturwissenschaft welche auf empirische Principien;

7 gegründet ist der **Physik**. <- Physik aber ist die Lehre von den **bewegenden Kräften** welche der

8 Materie eigen sind >

9 Die mathematische *mathematik* (wie sie Newton in seinem unsterblichen Werke vor-

10 trägt) haben die letztere nicht zum **Gegenstände**, nämlich nicht die vorangehende

11 **bewegende Kräfte**, welche ohnedem *1* durch Erfahrung müßen erkannt werden

12 sondern nur die Gesetze der **Bewegung** z. B. daß ein Körper *2* der durch zwey

13 **Kräfte** nach zwey Richtungen die *3* einen Winkel einschließen zugleich getrieben

14 **we** **bewegt** wird die Diagonallinie eines Parallelograms in derselben Zeit

15 **beschreibe** in welcher er jede der Seiten besonders durchlaufen haben würde, u. d. g.

16 und weñ in Ansehung der Kreisbewegung von Centrifugal- und Centri-

14 Es ist aber unmöglich aus bloß empirischen Erkenntni Begriffen

15 ein System zu zīmern. Es wird jederzeit ein zusāmen gestoppeltes zusāmengestoppeltes

16 Aggregat bleiben von Beobachtungen dieser oder jener Eigenschaft der

17 Materie bleiben was zwar ansehnlich aber doch īmēr nur fragmen-

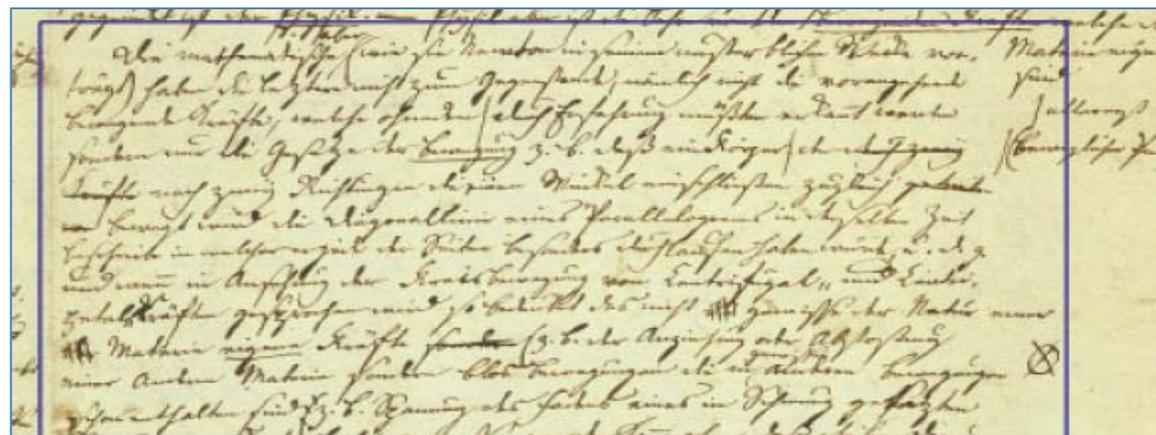
18 tarisch wachsen kān und **5* wē man still stehen kān wo man will

19 weil es an der Idee fehlt mangelt welche ein innerlich #begründetes

20 und zugleich sich selbst #begrenzendes Gantze ausmacht; ein solches aber

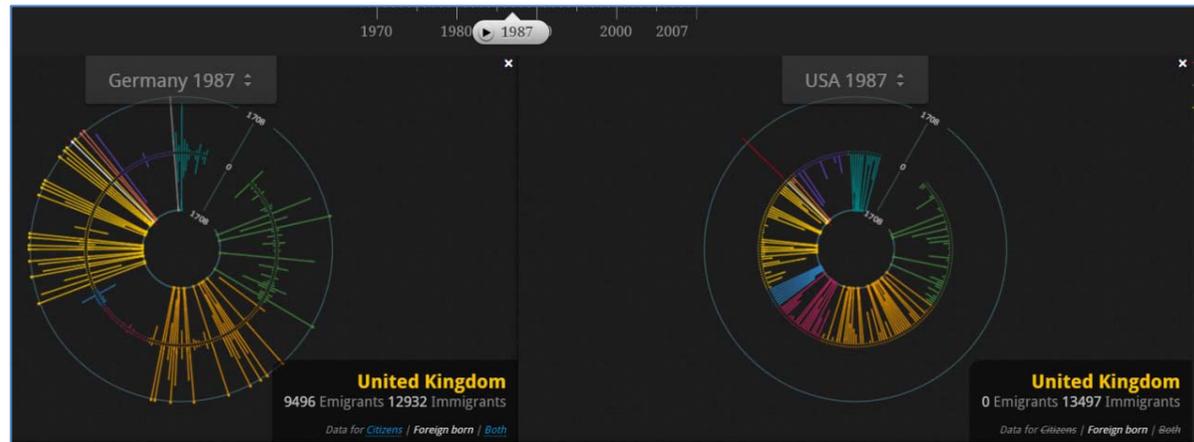
21 also kān nicht anders als nach Begriffen *a priori* zu Stande kom-

22 men.



Präsentationsschicht

- Visualisierung mit ProcessingJS
- Ausgabe für den menschlichen Nutzer in einer Browserumgebung
- Interaktionsmöglichkeiten mit Faceted Browsing via JavaScript



Middleware

- Abfrage und Zusammenstellung der Primärdaten mittels einer Ruby-on-Rails-Webapplikation
- Weitergabe im JSON-Format an die Client-Anwendung im Browser



```
application_controller.rb  entries_controller.rb  flow (git: master)
55
56 # GET /entries/1/edit
57 def edit
58   @entry = Entry.find(params[:id])
59 end
60
61 # POST /entries
62 # POST /entries.xml
63 def create
64   @entry = Entry.new(params[:entry])
65
66   respond_to do |format|
67     if @entry.save
68       format.html { redirect_to(@entry, :notice => 'Entry was successfully created.' ) }
69       format.xml { render :xml => @entry, :status => :created, :location => @entry }
70     else
71       format.html { render :action => "new" }
72       format.xml { render :xml => @entry.errors, :status => :unprocessable_entity }
73     end
74   end
75 end
```

Primärdaten

- Bereitstellung in einer Datenbank (bspw. MySQL)



id	scope_id	scope_iso_code	country_id	country_iso_code	country_name	continent_id	continent	year	emigrants_both	emigrants_c
324397	549	DEU	534	CHN	China	NUL	Asia	1996	4961	
324398	549	DEU	534	CHN	China	NUL	Asia	1997	5578	
324399	549	DEU	534	CHN	China	NUL	Asia	1998	5923	
324400	549	DEU	534	CHN	China	NUL	Asia	1999	5719	
324401	549	DEU	534	CHN	China	NUL	Asia	2000	6290	
324402	549	DEU	534	CHN	China	NUL	Asia	2001	6826	
324403	549	DEU	534	CHN	China	NUL	Asia	2002	9459	
324404	549	DEU	534	CHN	China	NUL	Asia	2003	11999	
324405	549	DEU	534	CHN	China	NUL	Asia	2004	13730	
324406	549	DEU	534	CHN	China	NUL	Asia	2005	11966	