

Forschungsdaten aus dem Archiv – und was Historiker*innen mit ihnen machen können

Thomas Wallnig (Wien)

Zur Problemstellung

Lena-Luise Stahn (Luhmann-Archiv): RNAB und digitale Ressourcen - was benötigt die Wissenschaft? Impulse aus den Digital Humanities (28. September 2023, Workshop “RNAB und digitale Ressourcen”, Deutsche Nationalbibliothek, Exilarchiv)

=> beschreibende Regelsysteme sollten sich nicht von vornherein ausschließen; dazu sollten sie in ihrer Modellierung Entitäten und Relationen sauberer trennen

=> digitale Repräsentation als eigenes zu verknüpfendes Objekt auffassen; das könnte auch für TEIs von Transkriptionen gelten

Beispiel Abschiedsvorlesung: [Audio-Aufzeichnung](#) sowie [Manuskript](#)

Zur Problemstellung

```
▼<TEI xmlns="http://www.tei-c.org/ns/1.0" xmlns:html="http://www.w3.org/1999/xhtml" xmlns:xi="http://www.w3.org/2001/XInclude"
  xmlns:xs="http://www.w3.org/2001/XMLSchema" rendition="luhmann_ms">
  ▼<teiHeader>
    ▼<fileDesc>
      ▼<titleStmnt>
        <title type="main">"Was ist der Fall?" und "Was steckt dahinter?"</title>
        <title type="sub">Die zwei Soziologien und die Gesellschaftstheorie</title>
        <title type="short"/>
      </titleStmnt>
      ▼<publicationStmnt>
        ▼<publisher>
          <ref target="https://niklas-luhmann-archiv.de">Niklas Luhmann-Archiv</ref>
        </publisher>
        <pubPlace>Bielefeld</pubPlace>
        <date>2020</date>
        ▼<availability status="free" rend="goobi-display portal-display">
          <licence target="https://creativecommons.org/licenses/by-nc-sa/4.0/">Creative Commons Attribution-ShareAlike 4.0 International (CC BY-
            NC-SA 4.0)</licence>
        </availability>
      </publicationStmnt>
      ▼<sourceDesc>
        ▼<msDesc xml:id="MS_1517" type="typescript" subtype="typewriter">
          ▼<msContents class="print">
            ▼<msItem>
              <idno type="Niklas-Luhmann-Werk" corresp="#nla_W_1145"/>
              <idno type="Niklas-Luhmann-Findbuch"/>
            </msItem>
            ▼<author>
              ▼<name key="luhmann_niklas">
                <forename>Niklas</forename>
                <surname>Luhmann</surname>
              </name>
            </author>
```

Zur Problemstellung

Aufgabenstellung aus dem [DH-Skills-I-Kurs](#) (nach ausführlichem Gebrauch von [Data Camp](#)):

Also upload half a page in which you present at least **two datasets** from two different repositories you consider working with in your final project. You should first provide a citation for the dataset, then write a sentence about what it represents, how it was generated (and by whom), what it contains, how it is structured, where it is stored and how it is licensed; and what steps would be necessary for you to work with it. Look at the Austrian DH repositories Phaidra, Aussda, Arche, Gams in the first place, but also at GoogleDatasets and [elsewhere](#).

[...]

Tweet political spheres / Macbeth; Insta-Dead / own dataset (scraped); APIS humanities scholars / Geodatabase shipwrecks; Graphic Narrative Corpus / MOMA Exhibition Dataset; Saga Corpus / Archive for Danish Literature; Survey of Scottish Witchcraft / Dialect Cultures; Collection data Carnegie Pittsburgh / Der Sturm / Visualizing geospatial data; Data.gv.at / London Data Store; Diosiris Ancient Greek Corpus / Opera Graeca Adnotata; Deutsches Exilarchiv: Exilpresse Digital / Artikel der jüdischen Periodika; Bechdel Test Dataset / Documenting the American South / Movie Script Database; Language use in Carinthia / Military Units in Yugoslav Wars / Music genres and genre mapping - own dataset; Global Trends in Mental Health Disorder / Netflix Movies / Popular Video Games; Darwin Letters / Freie Netzpublikationen; MiCREATE - Migrant Children and Communities / Refugee Experience in Alan Gratz's Refugee and Gillian Cross' After Tomorrow; American Presidency Project / Twitter Parliamentarian / Clinton & Trump Tweets; Olympic Historical Dataset / Graphic Narrative Corpus; Salem Witchcraft / Survey of Scottish Witchcraft

Zur Problemstellung

 [Product](#) [Solutions](#) [Open Source](#) [Pricing](#)

 just4jc / **DataCamp-3** Public

forked from [Anthonymcqueen21/DataCamp](#)

[Code](#) [Pull requests](#) 1 [Actions](#) [Projects](#) [Wiki](#) [Security](#) [Insights](#)

 **Files**

 master

Go to file

> 00-certificates

> 01-intro-to-python-for-data-scie...

> 02-intermediate-python-for-data...

> 1-matplotlib

> 2-dictionaries-and-pandas

> 3-logic-control-flow-and-filtering

> 4-loops

 _chapter-details.png

 add-column-(1).py

 add-column-(2).py

DataCamp-3 / **02-intermediate-python-for-data-science** / **4-loops** / **brics.csv** 

 **Blake Cannon** massive cleanup

Preview **Code** **Blame** 6 lines (6 loc) • 187 Bytes

Search this file

1		country	capital	area	population
2	BR	Brazil	Brasilia	8.516	200.4
3	RU	Russia	Moscow	17.10	143.5
4	IN	India	New Delhi	3.286	1252
5	CH	China	Beijing	9.597	1357
6	SA	South Africa	Pretoria	1.221	52.98

Zur Problemstellung

Aufgabenstellung aus dem [DH-Skills-I-Kurs](#) (nach ausführlichem Gebrauch von [Data Camp](#)):

Also upload half a page in which you present at least **two datasets** from two different repositories you consider working with in your final project. You should first provide a citation for the dataset, then write a sentence about what it represents, how it was generated (and by whom), what it contains, how it is structured, where it is stored and how it is licensed; and what steps would be necessary for you to work with it. Look at the Austrian DH repositories Phaidra, Aussda, Arche, Gams in the first place, but also at GoogleDatasets and [elsewhere](#).

[...]

Tweet political spheres / Macbeth; Insta-Dead / own dataset (scraped); APIS humanities scholars / Geodatabase shipwrecks; Graphic Narrative Corpus / MOMA Exhibition Dataset; Saga Corpus / Archive for Danish Literature; Survey of Scottish Witchcraft / Dialect Cultures; Collection data Carnegie Pittsburgh / Der Sturm / Visualizing geospatial data; Data.gv.at / London Data Store; Diosiris Ancient Greek Corpus / Opera Graeca Adnotata; Deutsches Exilarchiv: Exilpresse Digital / Artikel der jüdischen Periodika; Bechdel Test Dataset / Documenting the American South / Movie Script Database; Language use in Carinthia / Military Units in Yugoslav Wars / Music genres and genre mapping - own dataset; Global Trends in Mental Health Disorder / Netflix Movies / Popular Video Games; Darwin Letters / Freie Netzpublikationen; MiCREATE - Migrant Children and Communities / Refugee Experience in Alan Gratz's Refugee and Gillian Cross' After Tomorrow; American Presidency Project / Twitter Parliamentarian / Clinton & Trump Tweets; Olympic Historical Dataset / Graphic Narrative Corpus; Salem Witchcraft / Survey of Scottish Witchcraft

Zur Problemstellung

TASK 2. We will now look at some of the smaller and larger states, based on real data drawn from a statistical survey (Höck) published in 1800. The city of Ulm had ca. 37.000 inhabitants, the margraviate of Baden-Durlach 26.000, the HabsburgLands 21 millions, Prussia 6 millions. Create a list named `HRR_population` in which you store these values (in thousands)! Print the third and fourth items of the list. Reverse the list and print it out. Comment what you do in a separate markdown field. Is this a line of code that will be part of your chunk: `np.mean[HRR_pop]`? If not, why? (10 points)

TASK 3. Ulm had a territory of 830 square kilometers, Baden-Durlach 1.631, HabsburgLands (plus Hungary) 600.000 and Prussia 316.000. Import numpy as `np` and create an array of `HRR_pop` and `HRR_area`. Then calculate and print out the density (population divided by area). Calculate the mean population and print it out. (10 points)

TASK 4. Let's have a closer look at Prussia's area development over time: 35.728 (square kilometers) in 1608, 80.826 in 1640, 109.830 in 1688, 117.928 in 1740 and 316.232 in 1804. Import Matplotlib's pyplot as `plt`, create two lists for the axes year and area, and then display the plot in the most appropriate way! Write a short comment on why you chose this type of visualization! (8 points)

TASK 5. Now we want to build a dictionary of dictionaries, called `hrr_dict`. It should contain the keys population, area and confession, and report the observations made earlier. In our list, only the Habsburg lands were Catholic. Print out the dictionary. Then import pandas as `pd` and transform the dictionary into a DataFrame (note the spelling!) and print it out. Now select and print the information about all states with less than 900 square kilometers area. Finally find which protestant states had an area of more than 900 square kilometers. - What will be wrong with this line of code: `HRR_dict_df.iloc["ulm"]`? (16 points)

Zur Problemstellung

STATISTISCHE ÜBERSICHT

DER

DEUTSCHEN STAATEN,

IN ANSEHUNG

IHRER GRÖSSE, BEVÖLKERUNG, PRODUCTE,
INDUSTRIE UND FINANZVERFASSUNG,

VON

JOHANN DANIEL ALBRECHT HÖCK,

KÖN. PREUSS. JUSTITZRATH UND POLIZEIDIRECTOR IN SCHWABACH.



BASEL,

gedruckt bei J. DECKER, Buchhändler,
und so haben

DARMSTADT, in der PREUSS. KÖNIGLICHEN BUCHHANDLUNG, LEIPZIG, bey F. A. LAS.

Namen der Länder.	Grösse in Q. Meilen.	Bevölk.	Bevölk. pro Q. Meile.	Produkte	Industrie	Handel	Finanzverfassung und Münzverfassung.
I. DAS KÖNIGREICH BOHMEN.							
1. Das Städt. Prag.	12	4	330	100,000	100,000	100,000	100,000
2. Der Kaiserliche Kreis.	12	4	330	100,000	100,000	100,000	100,000
3. — Böhmen —	12	4	330	100,000	100,000	100,000	100,000
4. — Mähren —	2	11	140	10,000	10,000	10,000	10,000
5. — Schleisien —	4	44	110	10,000	10,000	10,000	10,000
6. — Ober-Oesterreich —	2	24	120	10,000	10,000	10,000	10,000
7. — Nieder-Oesterreich —	8	11	140	10,000	10,000	10,000	10,000
8. — Salzburg —	11	16	170	10,000	10,000	10,000	10,000
9. — Tirol —	8	16	200	10,000	10,000	10,000	10,000
10. — Steyermark —	12	16	200	10,000	10,000	10,000	10,000
11. — Kärnten —	8	11	140	10,000	10,000	10,000	10,000
12. — Friaul —	11	11	120	10,000	10,000	10,000	10,000
13. — Krain —	10	8	80	10,000	10,000	10,000	10,000
14. — Carinthien —	10	8	80	10,000	10,000	10,000	10,000
15. — Görz —	10	8	80	10,000	10,000	10,000	10,000
16. — Triest —	10	8	80	10,000	10,000	10,000	10,000
17. — Istrien —	10	8	80	10,000	10,000	10,000	10,000
18. — Venetien —	10	8	80	10,000	10,000	10,000	10,000
19. — Friaul —	10	8	80	10,000	10,000	10,000	10,000
20. — Krain —	10	8	80	10,000	10,000	10,000	10,000
21. — Carinthien —	10	8	80	10,000	10,000	10,000	10,000
22. — Görz —	10	8	80	10,000	10,000	10,000	10,000
23. — Triest —	10	8	80	10,000	10,000	10,000	10,000
24. — Istrien —	10	8	80	10,000	10,000	10,000	10,000
25. — Venetien —	10	8	80	10,000	10,000	10,000	10,000
26. — Friaul —	10	8	80	10,000	10,000	10,000	10,000
27. — Krain —	10	8	80	10,000	10,000	10,000	10,000
28. — Carinthien —	10	8	80	10,000	10,000	10,000	10,000
29. — Görz —	10	8	80	10,000	10,000	10,000	10,000
30. — Triest —	10	8	80	10,000	10,000	10,000	10,000
31. — Istrien —	10	8	80	10,000	10,000	10,000	10,000
32. — Venetien —	10	8	80	10,000	10,000	10,000	10,000
33. — Friaul —	10	8	80	10,000	10,000	10,000	10,000
34. — Krain —	10	8	80	10,000	10,000	10,000	10,000
35. — Carinthien —	10	8	80	10,000	10,000	10,000	10,000
36. — Görz —	10	8	80	10,000	10,000	10,000	10,000
37. — Triest —	10	8	80	10,000	10,000	10,000	10,000
38. — Istrien —	10	8	80	10,000	10,000	10,000	10,000
39. — Venetien —	10	8	80	10,000	10,000	10,000	10,000
40. — Friaul —	10	8	80	10,000	10,000	10,000	10,000
41. — Krain —	10	8	80	10,000	10,000	10,000	10,000
42. — Carinthien —	10	8	80	10,000	10,000	10,000	10,000
43. — Görz —	10	8	80	10,000	10,000	10,000	10,000
44. — Triest —	10	8	80	10,000	10,000	10,000	10,000
45. — Istrien —	10	8	80	10,000	10,000	10,000	10,000
46. — Venetien —	10	8	80	10,000	10,000	10,000	10,000
47. — Friaul —	10	8	80	10,000	10,000	10,000	10,000
48. — Krain —	10	8	80	10,000	10,000	10,000	10,000
49. — Carinthien —	10	8	80	10,000	10,000	10,000	10,000
50. — Görz —	10	8	80	10,000	10,000	10,000	10,000
51. — Triest —	10	8	80	10,000	10,000	10,000	10,000
52. — Istrien —	10	8	80	10,000	10,000	10,000	10,000
53. — Venetien —	10	8	80	10,000	10,000	10,000	10,000
54. — Friaul —	10	8	80	10,000	10,000	10,000	10,000
55. — Krain —	10	8	80	10,000	10,000	10,000	10,000
56. — Carinthien —	10	8	80	10,000	10,000	10,000	10,000
57. — Görz —	10	8	80	10,000	10,000	10,000	10,000
58. — Triest —	10	8	80	10,000	10,000	10,000	10,000
59. — Istrien —	10	8	80	10,000	10,000	10,000	10,000
60. — Venetien —	10	8	80	10,000	10,000	10,000	10,000
61. — Friaul —	10	8	80	10,000	10,000	10,000	10,000
62. — Krain —	10	8	80	10,000	10,000	10,000	10,000
63. — Carinthien —	10	8	80	10,000	10,000	10,000	10,000
64. — Görz —	10	8	80	10,000	10,000	10,000	10,000
65. — Triest —	10	8	80	10,000	10,000	10,000	10,000
66. — Istrien —	10	8	80	10,000	10,000	10,000	10,000
67. — Venetien —	10	8	80	10,000	10,000	10,000	10,000
68. — Friaul —	10	8	80	10,000	10,000	10,000	10,000
69. — Krain —	10	8	80	10,000	10,000	10,000	10,000
70. — Carinthien —	10	8	80	10,000	10,000	10,000	10,000
71. — Görz —	10	8	80	10,000	10,000	10,000	10,000
72. — Triest —	10	8	80	10,000	10,000	10,000	10,000
73. — Istrien —	10	8	80	10,000	10,000	10,000	10,000
74. — Venetien —	10	8	80	10,000	10,000	10,000	10,000
75. — Friaul —	10	8	80	10,000	10,000	10,000	10,000
76. — Krain —	10	8	80	10,000	10,000	10,000	10,000
77. — Carinthien —	10	8	80	10,000	10,000	10,000	10,000
78. — Görz —	10	8	80	10,000	10,000	10,000	10,000
79. — Triest —	10	8	80	10,000	10,000	10,000	10,000
80. — Istrien —	10	8	80	10,000	10,000	10,000	10,000
81. — Venetien —	10	8	80	10,000	10,000	10,000	10,000
82. — Friaul —	10	8	80	10,000	10,000	10,000	10,000
83. — Krain —	10	8	80	10,000	10,000	10,000	10,000
84. — Carinthien —	10	8	80	10,000	10,000	10,000	10,000
85. — Görz —	10	8	80	10,000	10,000	10,000	10,000
86. — Triest —	10	8	80	10,000	10,000	10,000	10,000
87. — Istrien —	10	8	80	10,000	10,000	10,000	10,000
88. — Venetien —	10	8	80	10,000	10,000	10,000	10,000
89. — Friaul —	10	8	80	10,000	10,000	10,000	10,000
90. — Krain —	10	8	80	10,000	10,000	10,000	10,000
91. — Carinthien —	10	8	80	10,000	10,000	10,000	10,000
92. — Görz —	10	8	80	10,000	10,000	10,000	10,000
93. — Triest —	10	8	80	10,000	10,000	10,000	10,000
94. — Istrien —	10	8	80	10,000	10,000	10,000	10,000
95. — Venetien —	10	8	80	10,000	10,000	10,000	10,000
96. — Friaul —	10	8	80	10,000	10,000	10,000	10,000
97. — Krain —	10	8	80	10,000	10,000	10,000	10,000
98. — Carinthien —	10	8	80	10,000	10,000	10,000	10,000
99. — Görz —	10	8	80	10,000	10,000	10,000	10,000
100. — Triest —	10	8	80	10,000	10,000	10,000	10,000
101. — Istrien —	10	8	80	10,000	10,000	10,000	10,000
102. — Venetien —	10	8	80	10,000	10,000	10,000	10,000
103. — Friaul —	10	8	80	10,000	10,000	10,000	10,000
104. — Krain —	10	8	80	10,000	10,000	10,000	10,000
105. — Carinthien —	10	8	80	10,000	10,000	10,000	10,000
106. — Görz —	10	8	80	10,000	10,000	10,000	10,000
107. — Triest —	10	8	80	10,000	10,000	10,000	10,000
108. — Istrien —	10	8	80	10,000	10,000	10,000	10,000
109. — Venetien —	10	8	80	10,000	10,000	10,000	10,000
110. — Friaul —	10	8	80	10,000	10,000	10,000	10,000
111. — Krain —	10	8	80	10,000	10,000	10,000	10,000
112. — Carinthien —	10	8	80	10,000	10,000	10,000	10,000
113. — Görz —	10	8	80	10,000	10,000	10,000	10,000
114. — Triest —	10	8	80	10,000	10,000	10,000	10,000
115. — Istrien —	10	8	80	10,000	10,000	10,000	10,000
116. — Venetien —	10	8	80	10,000	10,000	10,000	10,000
117. — Friaul —	10	8	80	10,000	10,000	10,000	10,000
118. — Krain —	10	8	80	10,000	10,000	10,000	10,000
119. — Carinthien —	10	8	80	10,000	10,000	10,000	10,000
120. — Görz —	10	8	80	10,000	10,000	10,000	10,000
121. — Triest —	10	8	80	10,000	10,000	10,000	10,000
122. — Istrien —	10	8	80	10,000	10,000	10,000	10,000
123. — Venetien —	10	8	80	10,000	10,000	10,000	10,000
124. — Friaul —	10	8	80	10,000	10,000	10,000	10,000
125. — Krain —	10	8	80	10,000	10,000	10,000	10,000
126. — Carinthien —	10	8	80	10,000	10,000	10,000	10,000
127. — Görz —	10	8	80	10,000	10,000	10,000	10,000
128. — Triest —	10	8	80	10,000	10,000	10,000	10,000
129. — Istrien —	10	8	80	10,000	10,000	10,000	10,000
130. — Venetien —	10	8	80	10,000	10,000	10,000	10,000
131. — Friaul —	10	8	80	10,000	10,000	10,000	10,000
132. — Krain —	10	8	80	10,000	10,000	10,000	10,000
133. — Carinthien —	10	8	80	10,000	10,000	10,000	10,000
134. — Görz —	10	8	80	10,000	10,000	10,000	10,000
135. — Triest —	10	8	80	10,000	10,000	10,000	10,000
136. — Istrien —	10	8	80	10,000	10,000	10,000	10,000
137. — Venetien —	10	8	80	10,000	10,000	10,000	10,000
138. — Friaul —	10	8	80	10,000	10,000	10,000	10,000
139. — Krain —	10	8	80	10,000	10,000	10,000	10,000
140. — Carinthien —	10	8	80	10,000	10,000	10,000	10,000
141. — Görz —	10	8	80	10,000	10,000	10,000	10,000
142. — Triest —	10	8	80	10,000	10,000	10,000	10,000
143. — Istrien —	10	8	80	10,000	10,000	10,000	10,000
144. — Venetien —	10	8	80	10,000	10,000	10,000	10,000
145. — Friaul —	10	8	80	10,000	10,000	10,000	10,000
146. — Krain —	10	8	80	10,000	10,000	10,000	10,000
147. — Carinthien —	10	8	80	10,000	10,000	10,000	10,000
148. — Görz —	10	8	80	10,000	10,000	10,000	10,000
149. — Triest —	10	8	80				

Zur Problemstellung

```
[10]: HRR_population = [27, 26, 21000, 6000]
      print(HRR_population[2:4])
```

```
[21000, 6000]
```

```
[11]: HRR_population.reverse()
      print(HRR_population)
```

```
[6000, 21000, 26, 27]
```

```
[12]: import numpy as np
      HRR_pop = np.array([27, 26, 21000, 6000])
      HRR_area = np.array([830, 1631, 600000, 300000])
      density = HRR_pop / HRR_area
      print(density)
      np.mean(HRR_pop)
```

```
[0.03253012 0.01594114 0.035      0.02      ]
```

```
[12]: 6763.25
```

```
[13]: HRR_dict = {"population":{"ulm":27, "baden-durlach":26, "habsburglands":21000, "prussia":6000}, "area":{"ulm":830, "baden-durlach":1631, "habsburglands":600000, "prussia":300000}
      HRR_dict
      import pandas as pd
      HRR_dict_df = pd.DataFrame(HRR_dict)
      print(HRR_dict_df)
      HRR_dict_df[HRR_dict_df["area"] < 900]
      np.logical_and(HRR_dict_df["area"] > 900, HRR_dict_df["confession"] == "protestant")
      HRR_dict_df.loc["ulm"]
```

	population	area	confession
ulm	27	830	protestant
baden-durlach	26	1631	protestant
habsburglands	21000	600000	catholic
prussia	6000	300000	protestant

```
[13]: population      27
      area            830
      confession      protestant
      Name: ulm, dtype: object
```

Zur Problemstellung

In jeder Art von geisteswissenschaftlicher Forschung werden digitale Ressourcen von Kulturerbe-Institutionen zur Auffindung von Objekten verwendet.

Digital betriebene Geisteswissenschaften benötigen darüber hinaus Forschungsdaten, die nachnutzbar sind. Nachnutzbar werden sie durch entsprechende technische Aufbereitung (Modellierung), adäquate Zurverfügungstellung (Download-Optionen, Schnittstellen) und entsprechende Lizenzierung.

Im Forschungsprozess angereicherte Daten sollten mit den ursprünglichen Datensätzen verknüpfbar gemacht werden.

Ebenso ist es hilfreich, wenn digitale Bearbeitungsprozesse in Kulturerbe-Institutionen gut dokumentiert werden. Hier ist die Grenze zu DH-Forschungsliteratur fließend.

Beispiele



https://gais.graz.at/stadtarchiv-graz/at/jr/fis/imdas/web/loadMask/view-mask-felder.jsf?objectId=893901&maskId=null&maskName=null

GRAZER ARCHIV INFORMATIONEN SYSTEM

Home Volltextsuche Suche Erweiterte Suche Hilfe

Bestandsstruktur

Letzte Detailsansicht

Informationen

Impressum

Kontakt

Die Kunstdenkmäler der Stadt Graz. Die Profanbauten des II., III. und VI. Bezirkes, bearb. von Erik Hilzensauer, 1. Aufl., Horn, Wien 2013. (Österreichische Kunsttopographie ; 60)

Die Kunstdenkmäler der Stadt Graz. Die Profanbauten des IV. und V. Bezirkes (Lend und Gries), bearb. von Amélie Sztatecsny; Horn, Wien 1984 . (Österreichische Kunsttopographie ; 46)

Eva Doppler: Virtuelle Rekonstruktion der zerstörten Synagoge in Graz, Wien, Techn. Univ., Dipl.-Arb., 2015.
https://publik.tuwien.ac.at/files/PubDat_241270.pdf

Wilhelm Steinböck (Hg.): Graz als Garnison. Beiträge zur Militärgeschichte der steirischen Landeshauptstadt, Graz, Wien: Leykam 1982.

Anmerkungen

Allgemeine Anmerkungen: Der Bestand wurde von Matthias Holzer und Gerhard Schwarz verzeichnet. Die Klassifikation, die Beschlagwortung und die Verknüpfung von Datensätzen mit Einträgen in der Gemeinsamen Normdatei (GND) wurde von Tamara Kefer durchgeführt.

Verzeichnungskontrolle

BearbeiterIn:	Tamara Kefer
Verzeichnungsgrundsätze:	ISAD(G)
Datum/Zeitraum der Verzeichnung:	15.12.2017
Link zur EAD-Datei (XML):	https://www.grazmuseum.at/gais/Plansammlung.xml
7.1 Status Bearbeitung:	in Bearbeitung
Lizenz Verzeichnungsdaten:	CC BY-NC 3.0 AT Creative Commons - Namensnennung-Nicht kommerziell 3.0 Österreich

Beispiele

Nachnutzbare Datenbestände der Österreichischen Nationalbibliothek, die man in der [Lehre](#) verwenden kann:

- [Botanische Illustrationen](#)
- [Historische Reiseberichte](#)
- [Bibliotheca Eugenia Digital](#)
- [Historische Postkartensammlung](#)
- [ÖNB-Katalog](#)
- [Digitale Editionen](#)
- [Wiener Zeitung](#)

Beispiele



**Landesarchiv
Baden-Württemberg**

FDMLab@LABW

Übersicht

Beiträge

Veröffentlichungen

Vorträge

Workshops

Projekt

Kontakt







Bild von Gerd Altmann auf Pixabay

Nach drei Jahren endet das Projekt FDMLab@LABW. Wir nutzen die Gelegenheit, um hier im Blog noch einmal einen Überblick über unsere Ergebnisse und Projekte zu geben.

 Einen ausführlichen [Rückblick über die Arbeit in den ersten beiden Jahren des FDMLab](#) → haben wir schon letztes Jahr geschrieben.

In der "Phase II" unserer Förderzeit haben wir uns mehr auf praktische Aufgaben konzentriert.

KI im Archiv

Das FDMLab wurde dazu eingeladen an den EDV-Tagen 2022 einen [Vortrag zum Thema KI im Archiv](#) → zu halten. Das Thema beschäftigte uns auch mit dem Aufkommen eines allgemeinen Zugangs zu Large Language Modellen via ChatGPT bei einer Gastvorlesung bei der [VU Digitalisierung an der Universität Wien im Wintersemester 2022/2023](#) ↗.

Gerade das Thema Künstliche Intelligenz ist ein Bereich, für den es scheinbar schon viele fertige Lösungen gibt. In der Domäne von archivischem Material liefern diese fertigen Lösungen häufig noch nicht die benötigte Qualität. Umso wichtiger ist ein regelmäßiger Austausch zu funktionierenden und nicht funktionierenden Ansätzen. Bei unserem Vortrag an den EDV-Tagen konnten wir nicht nur über unsere Erfahrungen berichten, sondern darauf aufbauend Feedback, Ideen und weitere Datenservices anderer Projekte kennen lernen.

OpenRefine

Die OpenRefine Workshops wurden mit weiteren Anleitungen versehen. Zum Beispiel, wie man [komplexe Datenabgleiche mit Wikidata](#) → und den [Getty Thesauri](#) → durchführen kann. Außerdem wurden weitere Tricks zum [Abgleich von Daten zwischen Projekten](#) → ergänzt.

Neben der Begleitung von Workshops, führte das FDMLab zusammen mit [Verena Mack](#) → von der [GND-Agentur LEO-BW-Regional](#) ↗ einen [Normdatenworkshop beim 82. Südwestdeutscher Archivtag](#) → durch.

 Die Inhalte der OpenRefine Workshops bleiben vorerst mit dem FDMLab Blog online verfügbar. Es gibt jedoch Überlegungen die Workshop Inhalte vom Blog zu entkoppeln und in einem anderen Format anzubieten

Auf dieser Seite

KI im Archiv

OpenRefine

Dokument-, Layout- und
Texterkennung

Provenienzforschung

Extraktion von
Schlagwörten zu Kunst- und
Kulturobjekten

Massenabgleich mit GND

Schlagwörter Netzwerk

NER für Metadaten im
Archiv

SpanCat für Personendaten

Weitere Datenprojekte

Reichskammergericht

Digitalisierung von
Heimlisten

Abgleich von Findbüchern
mit Dateistrukturen

Verlinkung der Tomi Actorum

Veröffentlichung von
Forschungsdaten

Schluss 



Beispiele

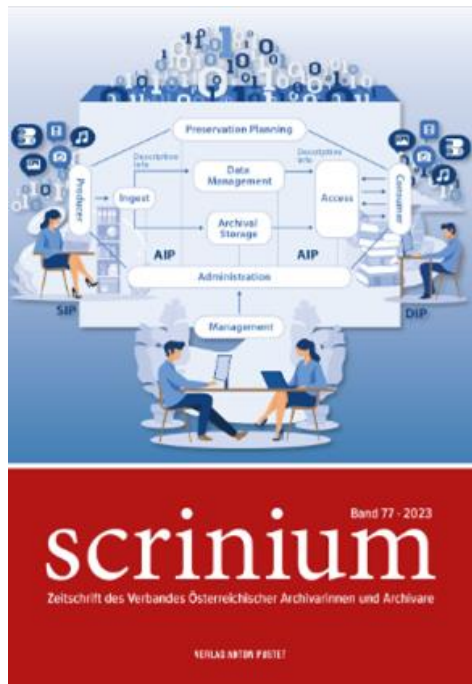
Konsequenzen der Handschriftenerkennung und des maschinellen Lernens für die Geschichtswissenschaft

Anwendung, Einordnung und Methodenkritik

von Tobias Model

I. Einleitung

Die Erkennung von Handschriften ist kein Kernthema der Geschichtswissenschaft, sondern eher eine wenig diskutierte Kompetenz von Historikerinnen und Historikern, die oft nicht nur gedruckte, sondern auch handschriftliche Dokumente auswerten. Wenn sie, in Archiv oder Bibliothek, nach Schriftstücken suchen, die sie für ihre Fragestellungen heranziehen können, und wenn sie diese Schriftstücke dann transkribieren, und sei es nur auszugsweise, investieren sie sehr viel Zeit. Von dieser Erfahrung ausgehend, scheint die maschinelle Erkennung von Text im Allgemeinen und Handschriften im Besonderen vor allem als zeitsparendes Hilfsmittel von Interesse für den Wissenschaftsbetrieb. Wechselt man die Perspektive und fragt, welche Bedeutung handschriftliches Material als empirische Grundlage historischer Forschung hat, kommt man zu einer anderen Einschätzung. Insbesondere wenn es um die Zeit vor 1900 geht, macht die handschriftliche Überlieferung einen großen Anteil der historischen Dokumente aus, die zu Quellen historischer Forschung werden können. Handschriftliche Dokumente computergestützt lesbar zu



THE MATTERHORN RDF DATA MODEL

Formalizing Archival Metadata With SHACL

Tobias Wildi
docuteam GmbH
Baden, Switzerland
t.wildi@docuteam.ch

Alain Dubois
Archives de l'Etat du Valais
Sion, Switzerland
alain.dubois@admin.vsi.ch

Matterhorn RDF is a linked data-based model for archival metadata with the goal of improving the contextualization of archival records. It covers the three standards ISAD(G), ISAAR(CPF) and ISDF, as well as the areas "Preservation Description Information" and "Representation Information" of the OAIS information model. For the implementation of Matterhorn RDF, classes and properties of existing ontologies are used. The formalization of the model is realized with the help of SHACL shapes. [1]

Keywords: Archival metadata model, linked data, ontology, SHACL, RIC, contextualization
Conference Topics: Exploring New Horizons.

I. INTRODUCTION

This paper describes a model for archival metadata based on semantic technologies. The model represents both descriptive and technical metadata, specifically the standards ISAD(G), ISAAR(CPF) and ISDF of the International Council on Archives (ICA), as well as "Representation Information" and "Preservation Description Information" from the OAIS information model. The model also takes into account the current work of the ICA's Expert Group on Archival Description (EGAD), but chooses a different design approach than their conceptual model Records in Context (RIC).

The first part of this document defines the goal and scope of Matterhorn RDF. The second part substantiates why semantic technologies are used for the model and how they eliminate the disadvantages of today's XML-based data models. The third part outlines the design principles of Matterhorn RDF. This includes the decision not to develop an new ontology but rather exclusively use classes and properties of existing ontologies. The Shapes

Constraint Language (SHACL) is used to formalize and validate Matterhorn RDF. The fourth and fifth parts explain the concept model and the class model of Matterhorn RDF. The most important and at the same time unspectacular finding of both these parts is the realization that the innovation of Matterhorn RDF lies in the adaptation of existing models and ontologies for use in archives. The last part provides an outlook on the potential of Matterhorn RDF in terms of its technical implementation.

II. IMPROVE CONTEXTUALIZATION AS A GOAL

Archival metadata have the function of keeping the context in which documents were created comprehensible over a long period of time. Archival material has to be placed in a context to have any value. Thus, documents are contextualised through the description of their content (What?), the actors involved (Who?) and the process of creation (How?). The triangle of what, who and how has been covered to date by the three standards ISAD(G), ISAAR(CPF) and ISDF. While EAD and EAC can be coded in XML, the same is not true for ISDF. The three standards were developed by ICA over several years, with the result that they partly overlap and it is now unclear as to how relationships between them are to be mapped. The aim of Matterhorn RDF is firstly to ensure the encoding of the three standards and secondly to show how relationships between them can be modelled.

The need to revise, standardize and improve the relationship between the existing standards also manifested itself within the ICA. The Expert Group on Archival Description (EGAD) was founded in 2012 with the task of developing a new model under the title "Records in Context". Matterhorn RDF is not to be seen as an alternative to RIC, but rather seeks to

SHORT
PAPER

18th International Conference on Digital Preservation
(PRES 2019, Amsterdam, The Netherlands)
Copyright held by the author(s). The text of this paper is published
under a CC BY-SA license (<https://creativecommons.org/licenses/by-sa/4.0/>).

1

PRES
2019

Ansätze für nachhaltige Konvergenzprozesse



[Home](#) [Products](#) ▾ [Advantages](#) [Services](#) ▾ [Support](#) [Clients](#) ▾ [About us](#) ▾ [News](#) [Contact](#) | [EN](#) ▾

[Home](#) > [Product overview](#) > [scopeArchiv](#)

Plug-Ins

For exporting and importing content (metadata and primary data)

Excel export plug-in

Export of data from the scopeArchiv modules to Excel.

Data import plug-in

Import of metadata to existing distortion units with the choice of data elements to be accutalized.

Word template plug-in

Export of data from the scopeArchiv modules to Word. Creation of inventories.

Files manager plug-in

Management of linked files of the distortion units like upload, download of selected files of selected distortion units, activation for publication via the QueryFilePublisher for Query.

EAD export plug-in

Export of metadata of the distortion unit according to the EAD standard APEX or DDB for the connection of the portals.

Services

Additional services

Ansätze für nachhaltige Konvergenzprozesse

Schritte in Richtung einer mittelfristigen Adressierung der genannten Anliegen

- strukturierte Zusammenarbeit in den einschlägigen Konsortien und Verbünden (z.B. clariah.at, DH-Infra...), incl. des Teilens und Aufnehmens von bereits bestehender Erfahrung & Literatur
- Zusammenarbeit in Lehrveranstaltungen / Erweiterung von bestehenden Lehrveranstaltungs-Syllabi um entsprechende Komponenten (Datenmodellierung, Normdaten, Records Management und historische Forschung, datenwissenschaftliche Analyseszenarien...)
- Design gemeinsamer Projektanträge (incl. Szenarien zur Verknüpfung von Forschungsdaten mit bestehender Infrastruktur)

Auf einer Meta-Ebene: systematisches und strukturiertes Bibliographieren und Durchführen von Desk Searches; inkrementelle und kritische Erhebungen des jeweiligen Wissensstandes