

Von der Handschrift zum Datensatz

Computergestützte Erschließung und
Aufbereitung historischer Wetterdaten

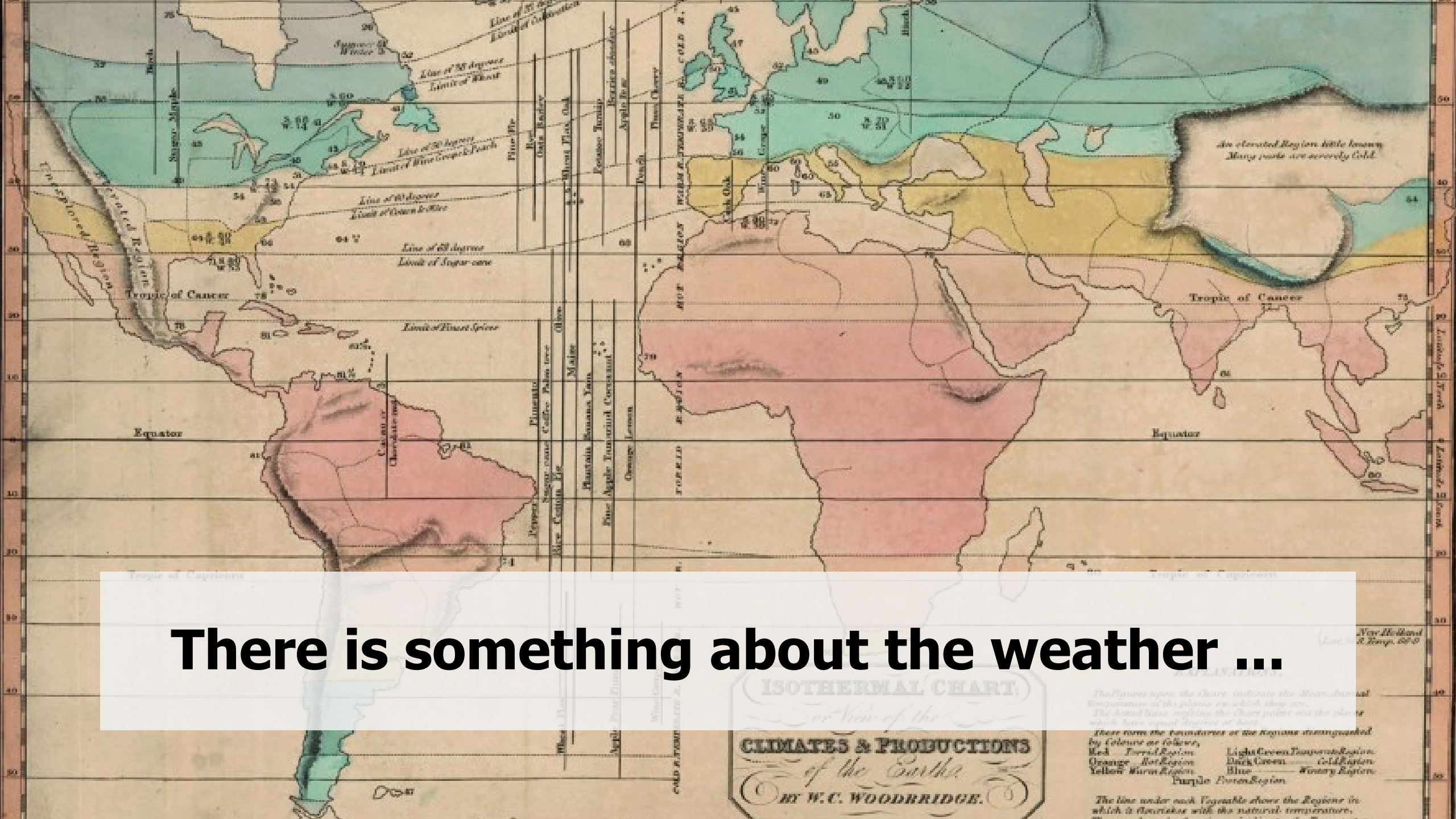
AG OCR | 08.07.2026

Constantin Lehenmeier (Universitätsbibliothek Regensburg)



AGENDA

- 1 *Observationes meteorologicae*: Historische Wetteraufzeichnungen aus Regensburg
- 2 Digitalisierungs- und Erschließungsprojekt der Wetteraufzeichnungen
- 3 Tabellen- und Texterkennung handschriftlicher Archivalien
- 4 Aufbereitung und Publikation der erschlossenen Daten
- 5 Bewertung und Fazit



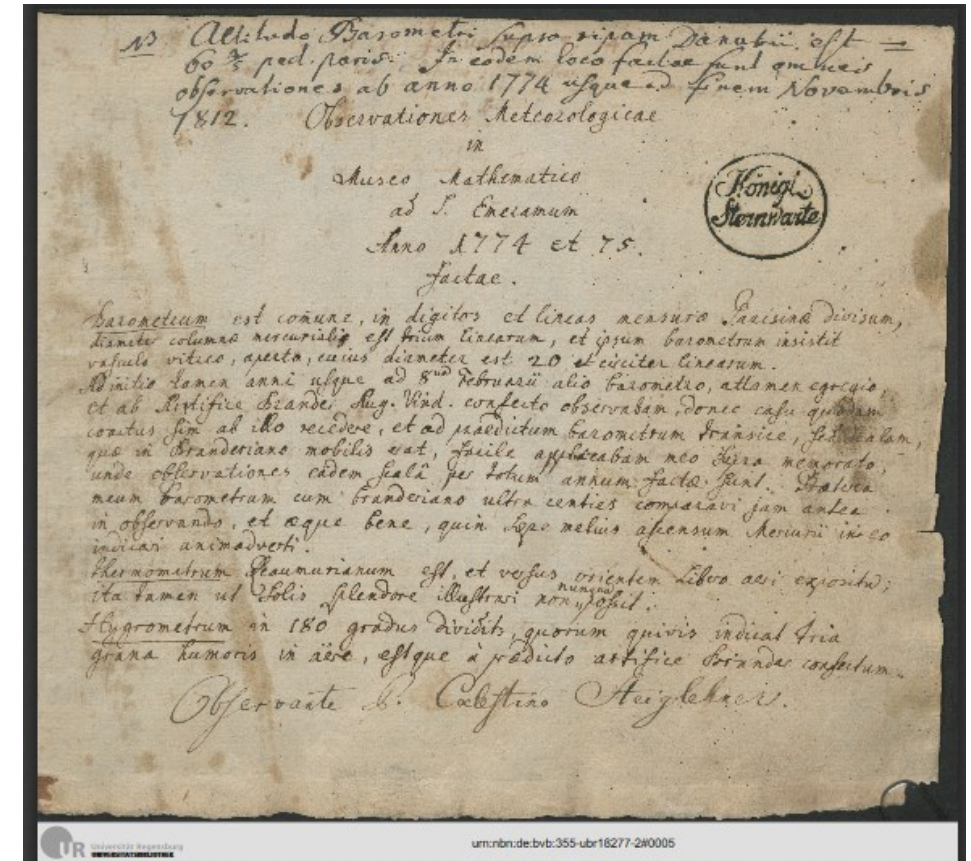
This is a detailed isothermal chart of the world, titled 'ISOTHERMAL CHART, or View of the CLIMATES & PRODUCTIONS of the Earth, BY W.C. WOODBRIDGE.' The map is color-coded into climate regions: Red for Torrid, Orange for Hot, Yellow for Warm, Purple for Temperate, Blue for Wintry, and Light Green for Frigid. Dotted lines represent isotherms, with numbers indicating the mean annual temperature in degrees Fahrenheit. The chart also shows the limits for various agricultural products, such as Sugar Maple, Pine Apple, Rice, and Wheat. The map includes latitude and longitude lines, and a legend at the bottom right explains the color coding and isotherm numbers.

There is something about the weather ...

Observationes meteorologicae

Mitunter die frühesten, so **kontinuierlich** geführten **Wetteraufzeichnungen** Europas¹:

- Angefertigt im Kloster St. Emmeram in Regensburg
- Handschriftliche Aufzeichnung von 1774 bis 1827
- Bis zu 16 verschiedene Messwerte pro Tag sowie Randnotizen und Kommentare



Titelseite des ersten erhaltenen Bandes „Observationes Meteorologicae in Museo Mathematico ad St. Emeramum Anno 1774 et 75 factae“

Forscher im Mönchsgewand



COELESTIN STEIGLEHNER
(1771-1779)



PLACIDUS HEINRICH
(1780-1825)



SCHÜLER
(1790-1802, 1804)



FERDINAND VON SCHMÖGER
(1824-1864)

Forscher im Mönchs

Zum Nutzen
für uns und
die Höheren!"



COELESTIN STEIGLEHNER
(1771-1779)



PLACIDUS HEINRICH
(1780-1825)



SCHÜLER
(1790-1802, 1804)



FERDINAND VON SCHMÖGER
(1824-1864)

Meteorologische Messdaten

Temperatur

Thermometer

innen	außen
Thermometer	
+4.0	-5.0
+3.5	-5.6
+3.4	-5.6

Luftdruck

Barometer, Manometer

Barom.	Manometer
27.0,5	+16 $\frac{1}{2}$
27.0.1	+17 $\frac{1}{2}$
27.0.0	-

Luftfeuchtigkeit

Lambert-Hygrometer,
Haarhygrometer

Hygrometer novio	velus 90 +
13.5	17 $\frac{1}{2}$
14.9	13 $\frac{3}{4}$
14.6	11 $\frac{1}{2}$
14.9	10
16.5	16 $\frac{1}{2}$

Windrichtung & -stärke

Windfahne

Ventus
SO $\frac{1}{2}$
SO 1
SO 1

Elektrische Ladung

Elektrometer

Declinatorium
18° 50'

Mehr als nur Messdaten

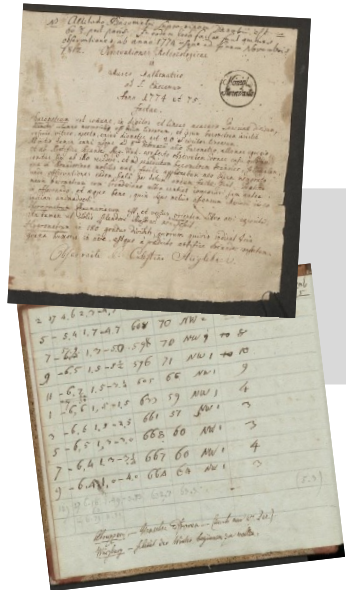
- Zustand der **Atmosphäre**, Sonnenfleckenbeobachtungen
- **Proxy-Daten**: Wasserstand und Vereisung der Donau, Saat-, Blüte und Erntezeiten
- Beschreibungen zu **Instrumenten und Geräten**
- **Naturphänomene** (Höhenrauch, Erdbeben San Francisco, Vesuv-Ausbruch)
- **Politisch und gesellschaftlich relevanten Ereignisse** als Randnotiz (Napoleon „die Stadt brennt“)

Sturm in London der einem Orkan gleich. Augst. Zeit. 1800 n° 34
Mit Blitz u. Donner.

Heute Abends Sturm zu Mannheim

Das heutige Donnerwetter, welches vorzüglich gegen Norden vorbezog, hat durch
Regenrissen und Hagel bey Kien - u. Hagel bey Deggendorf großen Schaden
gemacht.
26. 8. 60 1/2 38 1/2 27 1/2 34 1/2

Überblick über das Vorhaben



Dokumenterkennung



Layout- und
Texterkennung

Metadaten

Annotation



Maschinenlesbare
Aufbereitung

Metadaten

Publikation



Präsentation &
Verfügbarmachung

Metadaten

Nutzung



Analyse &
Benutzung

Dokumentation



Forschungsthemen

Thema 1: Modellierung historischer Tabellen als interoperable Datenstrukturen

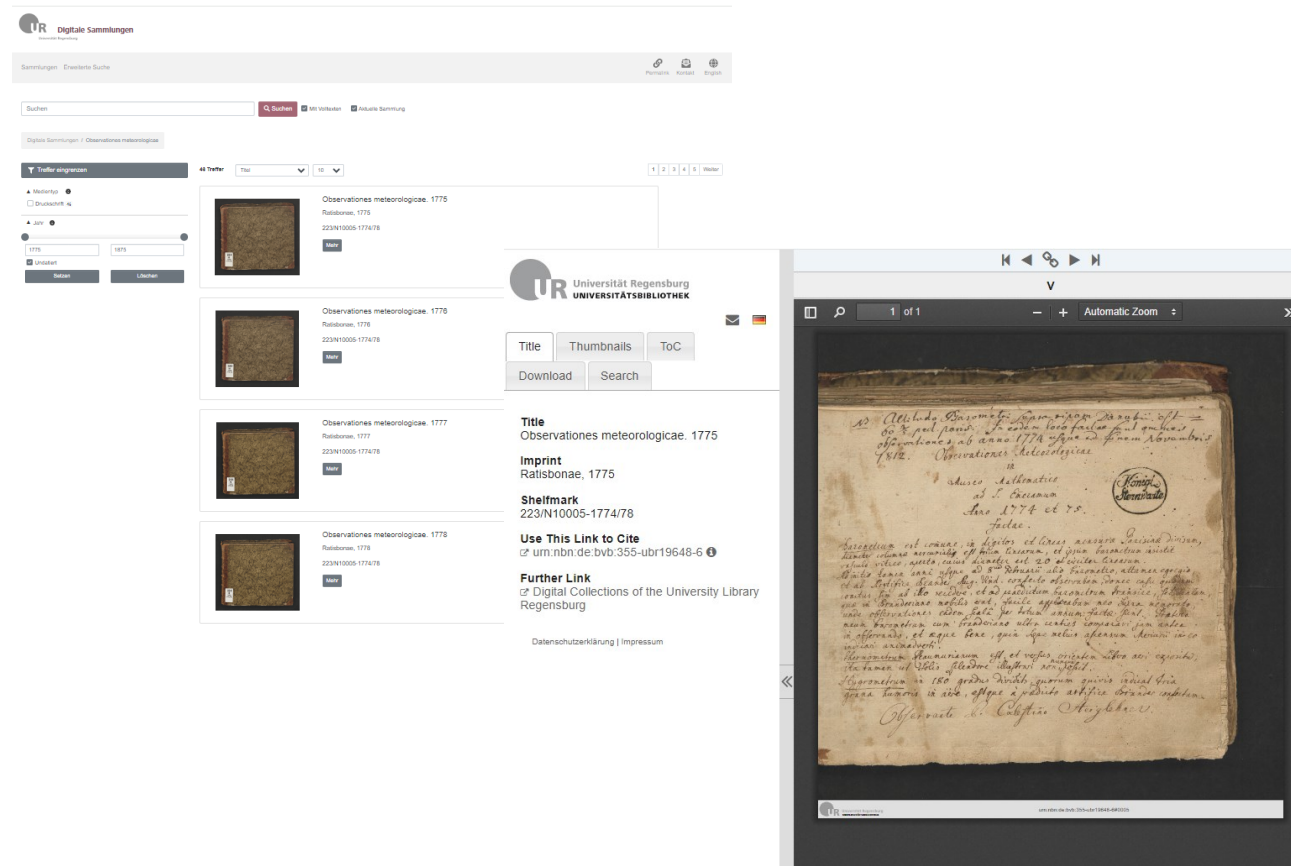
Thema 2: Algorithmische Rekonstruktion historischer Tabellenstrukturen zur Generierung nachnutzbarer Datensätze

Thema 3: Modell- und Datenkritik als Erweiterung historischer Quellenkritik

Digitalisierung an der Universitätsbibliothek Regensburg

Digitalisierung der vorhandenen Original-Bände

- 40 Bände von 1774 bis 1827
- 15.552 Seiten
- 10 Bände fehlen



Screenshot der digitalen Sammlung auf der Homepage der UB Regensburg

Überblick über die digitalisierte Sammlung

Seitenkategorie	Reine Tabellen	Tabellen & Text	Überwiegend Text	Mischform	Sonstiges (leer, Buchdeckel)
Anzahl $n = 200$	137	37	9	4	13
Prozentualer Anteil	68,5 %	18,5 %	4,5 %	2 %	6,5 %

Beispiele für „Seitenkategorien“ aus den Aufzeichnungen

„Warum nicht Transkribus?“

10. C	1.	26. 11. 5.	14. 9.	13. 2.	14. 2.				
6.	26. 11. 8.	14. 2.	11. 7.	84.	87.				
8.	27. 0. 2.	14. 8.	13. 8.	87.	87.				
10.	27. 0. 4.	15. 5.	16. 2.	21.	11.	V.			
1. 1/4	27. 0. 5.	15. 8.	16. 4.	47.	1.				
2 3/4	27. 0. 5.	15. 8.	15. 8.	59.	1.				
6 1/2	27. 0. 5.	14. 8.	12. 9.	47.	1.				
7	27. 0. 5.	15. 7. 14.	19. 10.	11. 5. 21. 4.					

- Geschlossenes System vs. Open-Source
- Lokale Verarbeitung soll möglich sein
- Tabellenspezifische Erkennung erforderlich

Erkannte Textzeilen und Tabellenzellen in *Transkribus*

Technischer Workflow

OCR-D Framework zu Verfahren von OCR

LAYOUTERKENNUNG

Regionen-Erkennung

- Seiten
- Tabellen
- Textregionen

Textzeilenerkennung innerhalb und außerhalb von Tabellen

↓
plus
Tabellenspalten

↓
Tabellenstruktur

TEXTERKENNUNG

Texterkennung mit
unterschiedlichen
Modellen

EXPORT

Export zu TEI und
CSV

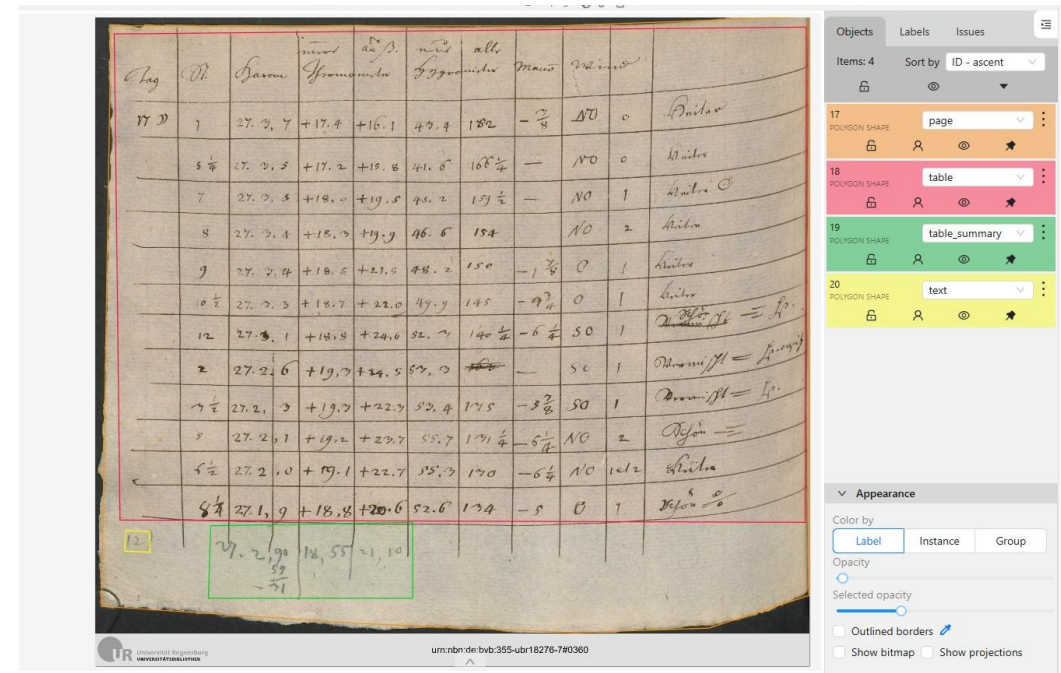
NER & Verlinkung

Layouterkennung

Integration einer YOLO-Erkennung in OCR-D

YOLO (*You-only-look-once*) zur effizienten Objekterkennung bei heterogenen Layouts *

- Wenig Hardware-Ressourcen notwendig und somit lokal effizient einsetzbar
- Schnelle Erstellung von Trainingsdaten für domänenspezifischen Korpus



* Kwon et al. (2022) zeigen, dass YOLOv5 in der Tabellenlokalisierung andere Modelle übertrifft; YOLO eingesetzt u.a. bei Clérice (2023)

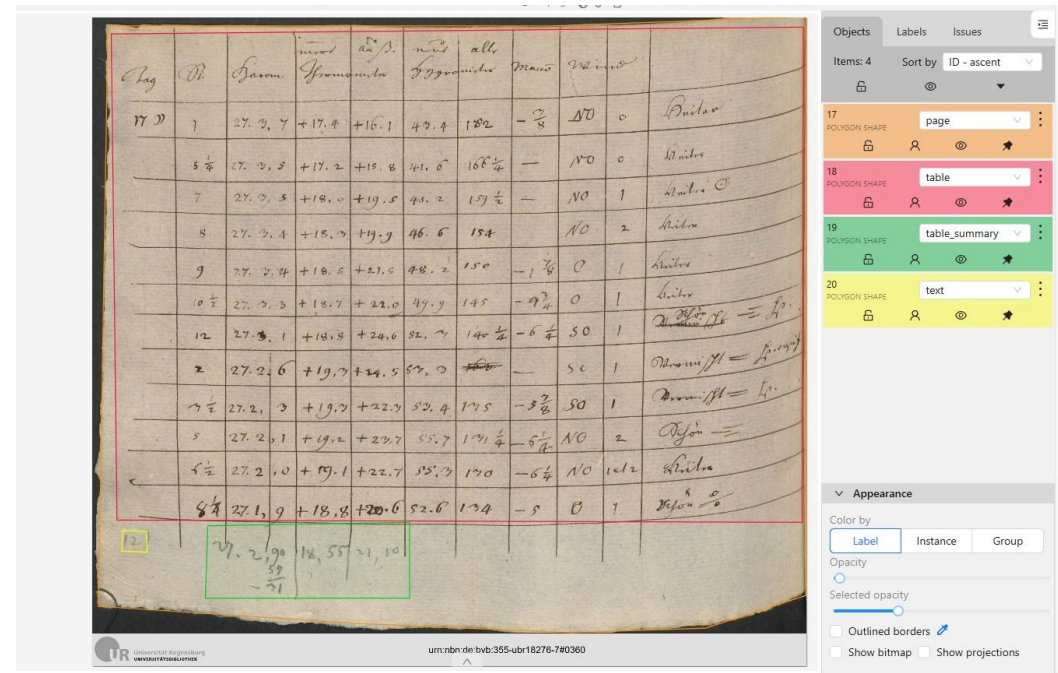
Layouterkennung

Integration einer YOLO-Erkennung in OCR-D

YOLO (*You-only-look-once*) zur effizienten Objekterkennung bei heterogenen Layouts *

- Wenig Hardware-Ressourcen notwendig und somit lokal effizient einsetzbar
- Schnelle Erstellung von Trainingsdaten für domänenspezifischen Korpus

Aktuell Evaluation zwischen Modellen von **YOLOv12** und **YOLOv26**



* Kwon et al. (2022) zeigen, dass YOLOv5 in der Tabellenlokalisierung andere Modelle übertrifft; YOLO eingesetzt u.a. bei Clérice (2023)

Layouterkennung

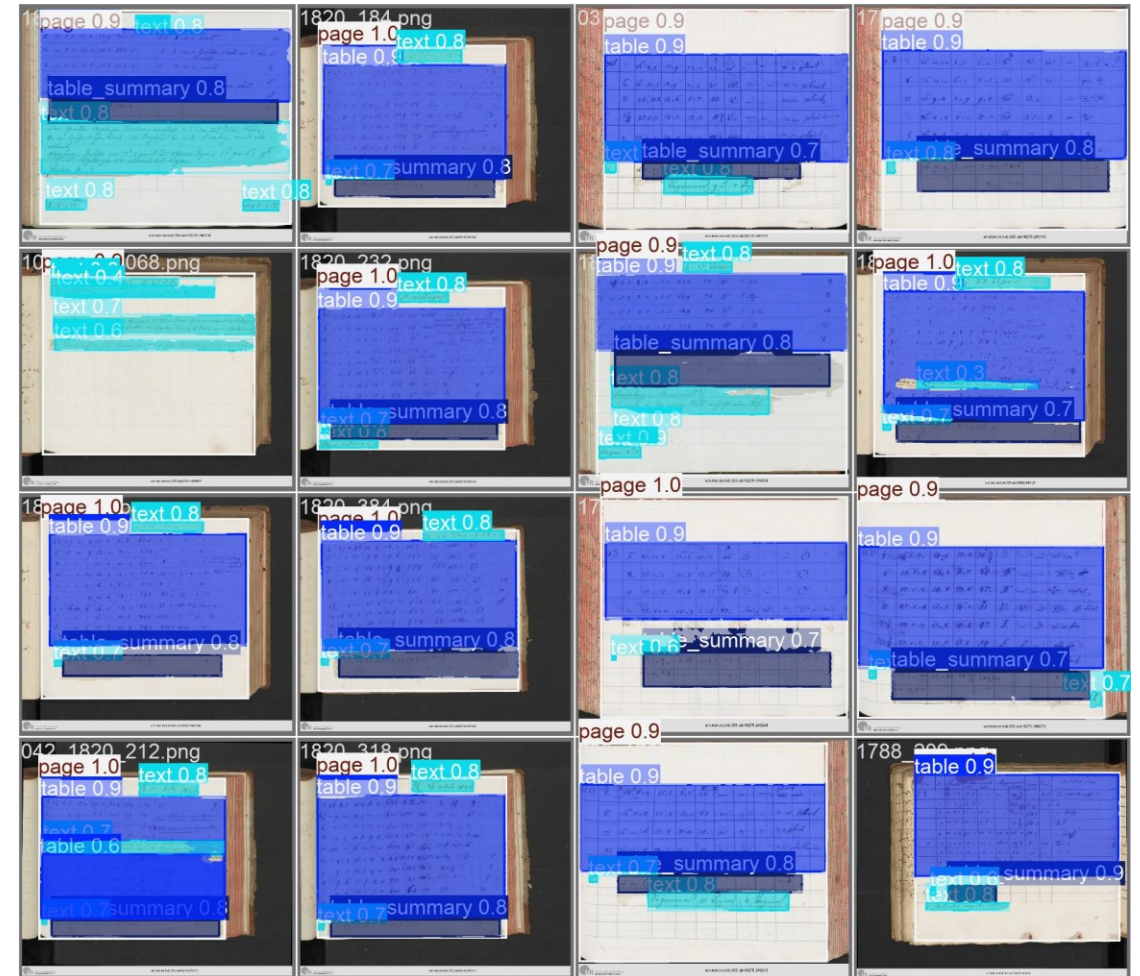
Training: 640

Validierung: 160

Test: 120

Elemente:

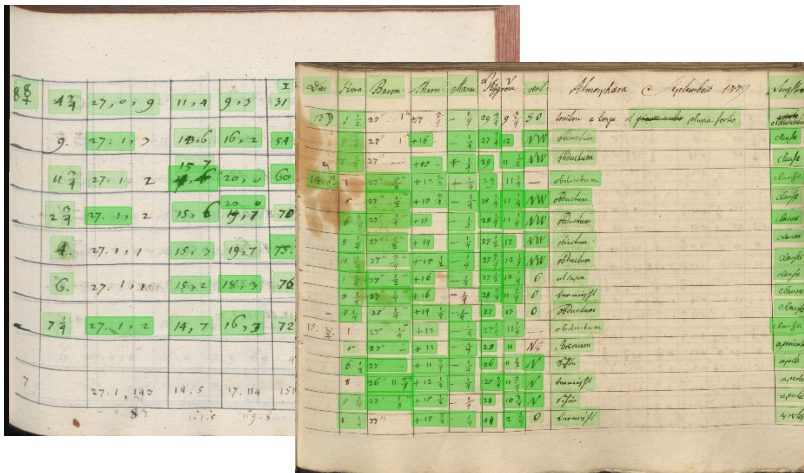
- Seitenrand ($MAP50-95$ 96%)
- Tabelle ($MAP50-95$ 95%)
- Textregion ($MAP50-95$ 89%)
- Tabelle (Mittelwerte) ($MAP50-95$ 91%)



Textzeilenerkennung

Innerhalb von Tabellen

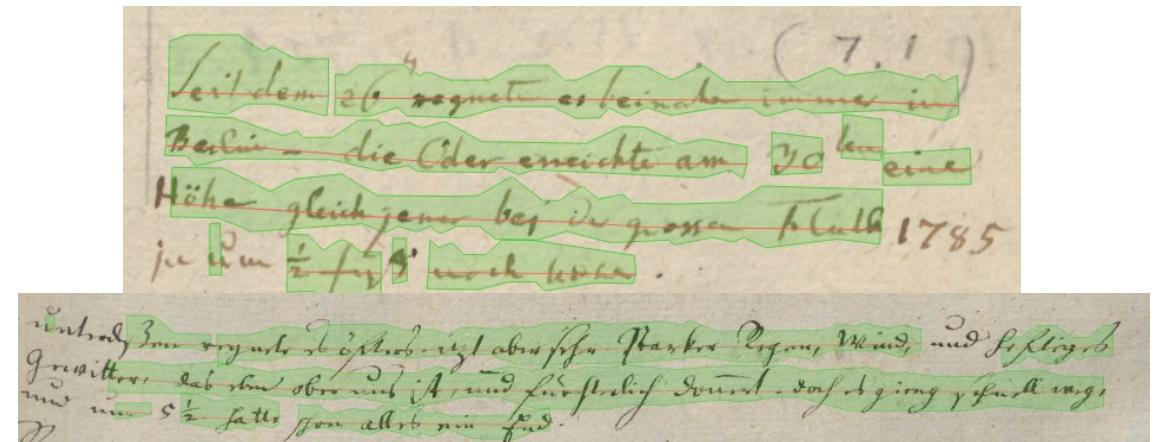
Eigenes **YOLO-Modell** trainiert und validiert mit 10.340 bzw. 1.690 Textzeilen (MAP_{50-95} 93%)



The image shows two pages from a handwritten table. The text is written in a cursive script. Green bounding boxes are drawn around individual text cells within the table structure. The table appears to be a ledger or a record book, with columns and rows of data. The text is written in a cursive script, likely from the 18th or 19th century.

Außerhalb von Tabellen

kraken¹



Spaltenerkennung

Training: 480

Validierung: 129

Test: 120

Elemente:

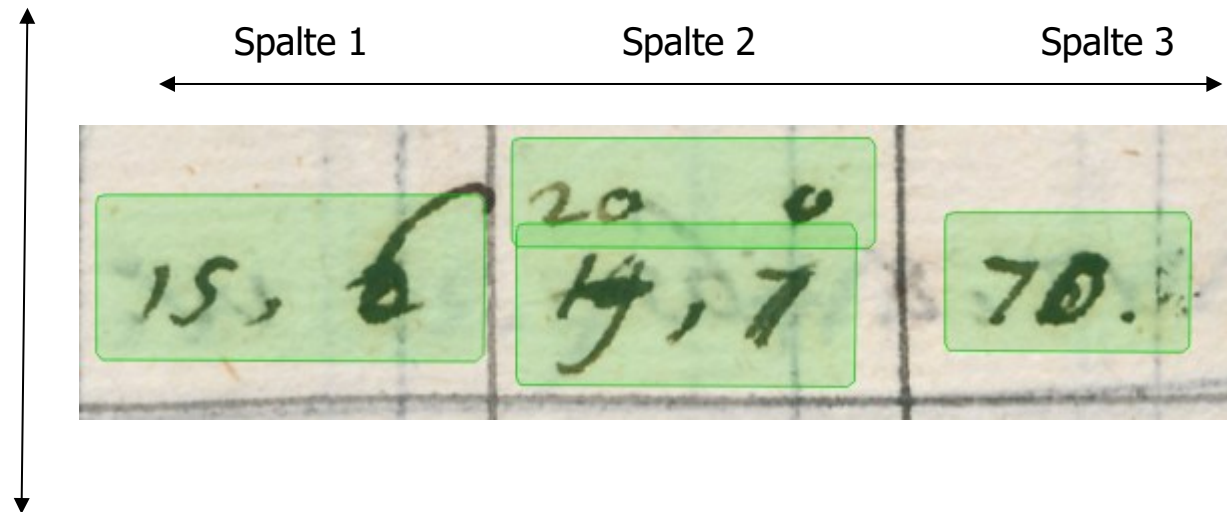
- Tabellenspalten (MAP_{50-95} 95%)
- Tabellenkopf (MAP_{50-95} 86%)

Thermometer	Höhe	Wind	Zeit	November
7.0.4	5.7	1.8	25.0	N. 14.45
7.0.16	5.4	1.2	19.9	
7.0.2	5.4	1.2	19.0	0. - 47
7.0.4	5.5	1.4	6.	0.
7.0.2	6.7	4.2	12.5	0.
7.0.3	7.4	5.9	30.0	0.
7.0.2	7.8	5.9	41.2	0. - 49
7.0.2	7.6	5.4	39.7	0.
7.0.3	6.5	2.4	22.4	0.
7.0.7	6.0	2.2	38.0	- 47 - 48

Tabellenerkennung

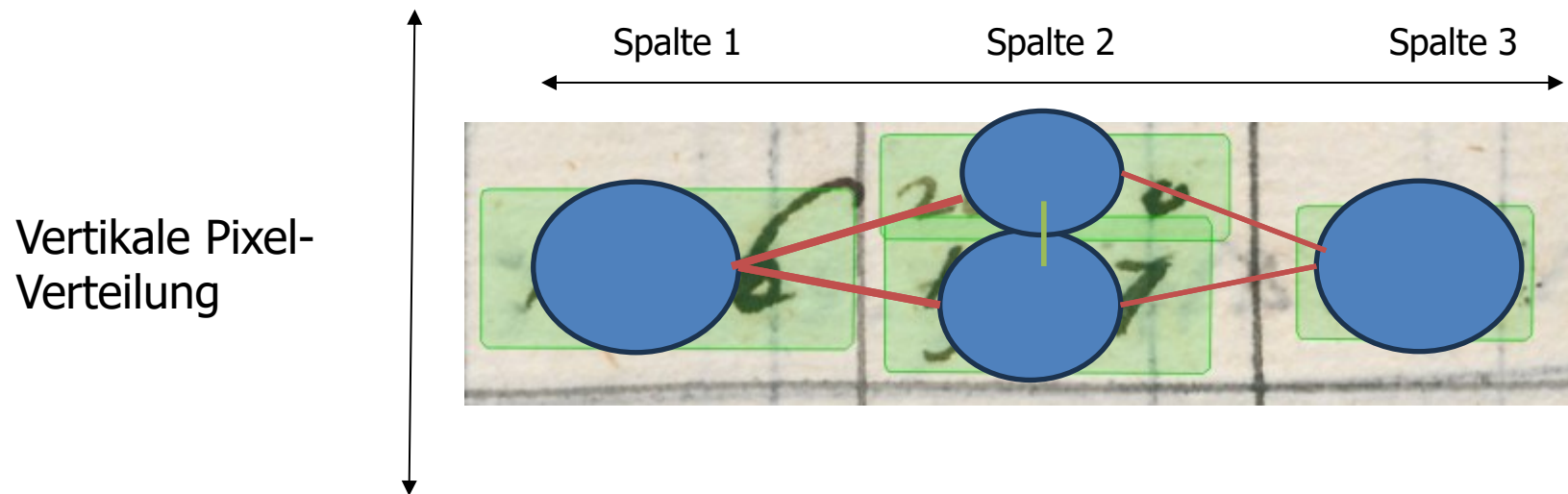
Clustering von Textzeilen zu Tabellenzeilen via KNN

Vertikale Pixel-
Verteilung



Tabellenerkennung

Clustering von Textzeilen zu Tabellenzeilen via KNN



Tabellenerkennung

Evaluation

4 18 July 1811

4 ¹	27.7.0	18.0	14.2	720	NW1	Er. l. m. a.
6	1.0	18.4	15.8	720	NW1	0
8	1.15	19.8	19.0	732	SO1	0
10	1.2	19.7	22.8	765	0	0
12	1.1	20.4	25.7	775	NO1	0 n. m. b. u. d.
2	1.0	20.6	25.9	781	SO2	—
5	0.9	21.2	26.6	790	NW1	0
6 ¹	1.0	20.8	24.6	794	NW1	0 n. m. b. u. d.
8	1.35	20.4	21.7	781	NW1	0 vel 0
10	1.7	20.1	18.7	754	N	0

10) 1.7 18 19.8 9 21.55 761 8

30 August 1811

5	27.2.2	14.2	7.6	608	NW1	0-8	5
7	2.6	18.0	8.8	663	NW1	— t. a.	7
9 ¹	2.7	15.3	10.8	700	NW1	— t. a.	7
11	2.7	15.5	12.2	745	NW1	— t. a.	6
14 ¹	2.6	15.6	12.0	778	NO1	— t. a.	5
24	2.6	15.6	12.5	785	NO1	— t. a.	5
5	2.6	15.3	12.7	786	NO1	—	7
7	2.7	15.0	10.6	778	NO1	0+	9
9 ¹	3.0	14.7	8.2	742	NO	0	10
10 ¹	3.2	14.5	7.5	740	NO	0	10

10) 27.2.695 15.07 10.35 7315

Rezen 4.7

Seitdem ist wegen der hohen ...
Boden die Oberseite am 30 ...
Höhe gleichsam bei der ...
in den ...

94% Strukturerkennungsrate (TEDS-S misst die strukturelle Ähnlichkeit zweier Tabellen)
n=40

Tabellenerkennung

Grenzen quantitativer Metriken

Herausforderung bei der Evaluation:
Wie gut eignen sich quantitative Metriken bei historischen Tabellen?

Wenig Guidelines sowie hohe Ambiguitäten erschweren die Ground-Truth-Erstellung

Tabellenerkennung

Grenzen quantitativer Metriken

Herausforderung bei der Evaluation:
Wie gut eignen sich quantitative Metriken bei historischen Tabellen?

Wenig Guidelines sowie hohe Ambiguitäten erschweren die Ground-Truth-Erstellung

NOTWENDIG

Quantitative Evaluation & qualitative Fehleranalyse

Dokumentation der Guidelines

Quantitative Metriken validieren Modelle, nicht historische Korrektheit (vgl. Weber, 2024)

Tabellenerkennung

Grenzen quantitativer Metriken

Media mensis Octobris valde incompleti.					
Die	Barom.	Th. int.	Th. ext.	Hygro	Summa
1					
2					
3					
4					
5					
6					
7	26. 10. 00.0	9. 470	7. 100	205. 7	1.7
8	7. 420	8. 920	6. 88	168. 2	1.5
9	9. 070	9. 070	7. 53	222. 1	1.0
10	13. 267	8. 800	7. 20	213. 1	5.7
11	375	10. 450	10. 675	219. 0	4.0
12	550	10. 700	10. 720	198. 0	2.25
13	970	10. 93	10. 700	191. 0	1.0
14	175	10. 90	10. 90	182. 4	2.0
15	13. 330	10. 87	9. 47	189. 0	1.0
16					
17					
18					
19	13. 100	9. 10	7. 325	194. 75	1.0
religiosa defunct.					
	26. 111, 457	99, 21	88, 06	1993, 85	21, 15
	26. 11, 1857	9, 921	8, 806	199, 385	2, 115
Summa					
	26. 11, 1957	Summa observationum 35.			

Texterkennung

Entwicklung weiterer OCR-D-Erweiterungen für zukünftige Evaluation in Frage kommender Texterkennungsmodellen:

*party*¹

Vortrainiertes Modell für Druck- und Handschriften

Mehrsprachigkeit, spezialisiert auf europäische Sprachen

vs

*Huggingface*²

Plattform zum Hosten verschiedenster Modelle, darunter zur Texterkennung

vs

Open LLMs

Lokale Nutzung durch API von LMStudio³

1: <https://github.com/mittagessen/party>
2: <https://huggingface.co/dh-unibe/tocr-kurrent>

3: <https://lmstudio.ai/>

Texterkennung

Aktuelle Ergebnisse

party

CER: 0.487

vs

Huggingface

TrOCR Modell „TrOCR Kurrent-
Model 19th century“

CER: 0.429

vs

Open LLMs

Qwen 3.5 9B

CER: 0.446

Gemma4 E4B

CER: 0.353

Texterkennung

Aktuelle Ergebnisse

party

CER: 0.487

vs

Huggingface

TrOCR Modell „TrOCR Kurrent-
Model 19th century“

CER: 0.429

vs

Open LLMs

Qwen 3.5 9B

CER: 0.446

Gemma4 E4B

CER: 0.353



Lokales Finetuning mit



unsloth

Annotation

PageXML

Repräsentation der visuellen
Layoutstruktur

Transkription durch
HTR



TEI, NER

Semantische
Annotation
Named-Entity Recognition
mit *hmBERT*¹

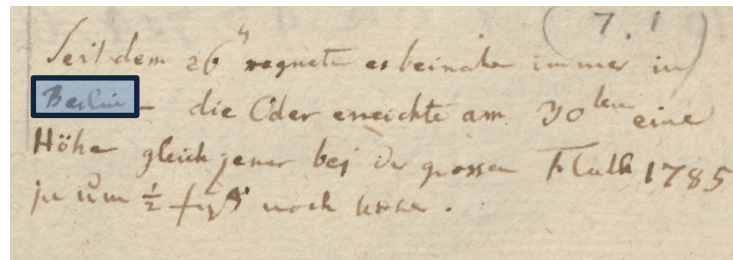


GND, Wikidata

Interoperabilität

Verknüpfung mit
Normdaten

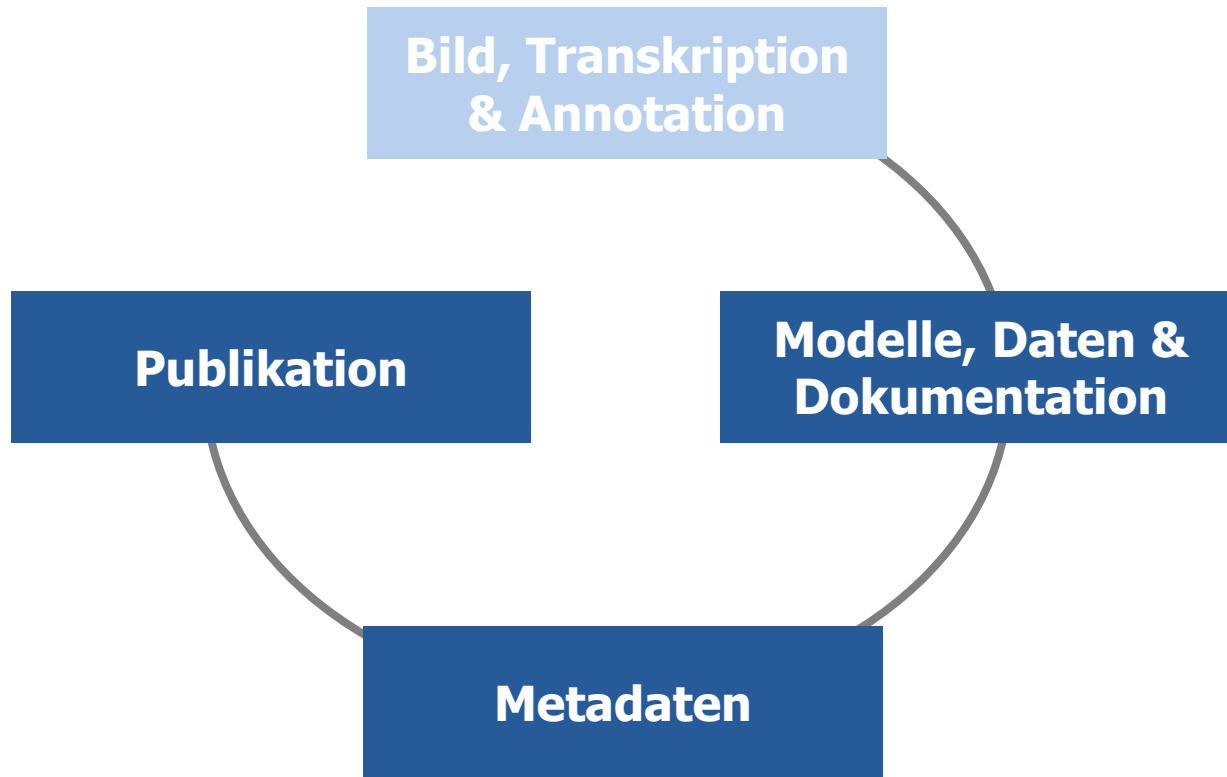
```
<pc:Textline id="region0004_TextRegion_line_3">
  <pc:Coords points="1450,2111 1450,2075"/>
  <pc:Baseline points="1450,2111 1568,2114 1776,2131 2130,2132"/>
  <pc:TextEquiv>
    <pc:Unicode>Berlin - die Oder erreichte am</pc:Unicode>
  </pc:TextEquiv>
</pc:Textline>
<pc:Textline id="region0004_TextRegion_line_4">
  <pc:Coords points="2159,2135 2158,2145 2217,2142 2218,2136 2218,2102 2159,2095"/>
  <pc:Baseline points="2159,2135 2218,2136"/>
  <pc:TextEquiv>
    <pc:Unicode>30</pc:Unicode>
  </pc:TextEquiv>
</pc:Textline>
<pc:Textline id="region0004_TextRegion_line_5">
```



Geografikum	Berlin
K10plusPPN:	10482638X 
GND-Nummer:	4005728-8 Link zu diesem Datensatz in der GND: http://d-nb.info/gnd/4005728-8
Katalogisierungsquelle:	rda
Typ:	Gebietskörperschaft, Verwaltungseinheit (gik); Gliedstaat (gif)
Quelle:	Orts-Mü. 28; B 2006 (unter 1.); M (unter 1.); Homepage (Stand: 28.05.2024): Link zur Quelle
Verwendungshinweis:	Ansetzung gilt für Veröffentlichungen der Stadt Berlin bis 1948, für gemeinsame Veröffentlichungen der vier Mächte für das durch den Vier-Mächte-Status definier Gesamt-Berlin und ab 3.10.1990
Weitere Angaben:	Hauptstadt und Bundesland; 1237 (Cölln) bzw. 1244 (Berlin) urkundl. erwähnt, 1230/40 bzw. nach 1250 Stadtrecht, 1307 Union, 1370 gemeinsamer Rat, 1432 Vereinigung der Doppelstadt

1: <https://huggingface.co/dbmdz/bert-base-historic-multilingual-cased>

FAIRe Wetterdaten



Finable

- DOI, Metadatenmodelle

Accessible

- HTTPS, OAI-PMH

Interoperabel

- Standardisierte und offene Formate (PageXML, TEI-XML), GND, Wikidata

Reusable

- Dokumentation, DOI, OA, Digitale Verfügbarmachung

Dokumentation der Modelle

MLFlow¹ zur Evaluierung und Monitoring der Modellerstellung

- Reproduzierbarkeit
 - Daten und Parameter werden dokumentiert
- Transparenz
 - Parameter und Logs können eingesehen werden
- Nachvollziehbarkeit
 - Modelle werden versioniert

1: <https://mlflow.org/>

Dokumentation der Modelle

MLFlow¹ zur Evaluierung und Monitoring der Modellerstellung

observations_yolo-seg-textline-dv1 ⓘ Provide Feedback ⓘ Add Description

Share

Runs Models Experimental Evaluation ⓘ Traces

⌵ ⌵ ⌵ 🔍 metrics.rmse < 1 and params.model = "tree" ⓘ Time created ▾ State: Active ▾ Datasets ▾ ⌵ Sort: Created ▾ ⌵ Columns ▾ ⋮ + New run

⌵ Group by ▾

	Run Name	Created ⌵	Duration	Source	Metrics				Parameters			
					metrics/mAP50-	metrics/mAP50↑	metrics/precisio	metrics/recallM	epochs	image_size	learning_rate	optimizer
☐	textline_detector_202604...	✓ 16 days ago	8.7h	train_yol...	0.57634164...	0.98250572...	0.94951010...	0.96108550...	180	1024	0.005	AdamW
☐	textline_detector_202604...	✗ 16 days ago	6.0s	train_yol...	-	-	-	-	180	1024	0.005	AdamW
☐	textline_detector_202604...	✓ 24 days ago	17.1h	train_yol...	0.59683402...	0.98316869...	0.95793041...	0.94674859...	180	1024	0.005	AdamW
☐	textline_detector_202604...	✗ 24 days ago	8.1s	train_yol...	-	-	-	-	180	1024	0.005	AdamW
☐	textline_detector_202604...	✗ 24 days ago	4.7s	train_yol...	-	-	-	-	180	1024	0.005	AdamW
☐	textline_detector_202604...	✗ 24 days ago	37.1s	train_yol...	-	-	-	-	180	1024	0.005	AdamW
☐	textline_detector_202604...	✗ 24 days ago	5.2s	train_yol...	-	-	-	-	180	1024	0.005	AdamW

1: <https://mlflow.org/>

Dokumentation der Modelle

MLFlow¹ zur Evaluierung und Monitoring der M

observations_yolo-seg-textline-dv1 [Provide Feedback](#) [Add Description](#)

Runs Models Experimental Evaluation [Traces](#)

[metrics.rmse < 1 and params.model = "tree"](#) [Time created](#) [State: Active](#) [Datasets](#) [Sort](#)

Group by

					Metrics		
☐	Run Name	Created ↕	Duration	Source	metrics/mAP50-	metrics/mAP50↑	metrics/precisio
☐	textline_detector_202604...	✓ 16 days ago	8.7h	train_yol...	0.57634164...	0.98250572...	0.94951010...
☐	textline_detector_202604...	✗ 16 days ago	6.0s	train_yol...	-	-	-
☐	textline_detector_202604...	✓ 24 days ago	17.1h	train_yol...	0.59683402...	0.98316869...	0.95793041...
☐	textline_detector_202604...	✗ 24 days ago	8.1s	train_yol...	-	-	-
☐	textline_detector_202604...	✗ 24 days ago	4.7s	train_yol...	-	-	-
☐	textline_detector_202604...	✗ 24 days ago	37.1s	train_yol...	-	-	-
☐	textline_detector_202604...	✗ 24 days ago	5.2s	train_yol...	-	-	-

observations_yolo-seg-textline-dv1 >

textline_detector_20260424_164805

Overview Model metrics System metrics Traces **Artifacts**

weights

BoxF1_curve.png
BoxPR_curve.png
BoxP_curve.png
BoxR_curve.png
MaskF1_curve.png
MaskPR_curve.png
MaskP_curve.png
MaskR_curve.png
args.yaml
confusion_matrix.png
confusion_matrix_normalized.png
labels.jpg
results.csv
results.png
train_batch0.jpg
train_batch1.jpg
train_batch17510.jpg
train_batch17511.jpg
train_batch17512.jpg
train_batch2.jpg
val_batch0_labels.jpg
val_batch0_pred.jpg
val_batch1_labels.jpg
val_batch1_pred.jpg
val_batch2_labels.jpg
val_batch2_pred.jpg

Share

ew run

imizer

imW

imW

imW

imW

imW

imW

imW

1: <https://mlflow.org/>

Dokumentation der Modelle



Metadatenschema für *Machine Learning*-Modelle

Aktuell: Erzeugung aus händisch
angelegter YAML-Datei



Zukünftig: *MLFlow* als Basis für die
Generierung

1: <https://github.com/RDA-FAIR4ML/FAIR4ML-schema>

Bereitstellung über Repositorium

Eigenes Repositorium basierend auf InvenioRDM¹

- Standardisiertes und anpassbares Metadatenformat
- Kontrolliertes Vokabular
- Versionierung
- Schnittstellen

The screenshot displays the InvenioRDM interface for a record titled '1813_248' by 'Lehenmeier, Constantin'. The interface is divided into several sections:

- Header:** Shows the record title '1813_248' and the author 'Lehenmeier, Constantin'.
- Files:** A table listing three files:

Name	Size	Download all
1813_248.jpg	1.8 MB	Preview Download
1813_248.xml	35.0 kB	Preview Download
1813_248_tei.xml	31.2 kB	Preview Download
- Additional details:**
 - Person:** Full Name (Placidus Heinrich), Birth Year (1756), Death Year (1825), Role (Primary Observer), Active Years (1774-1825).
 - Weather Observation:** Page Number (248), Transcription Status (In Progress), Date Start (1813), Volume (1813), Mentioned Places (Berlin).
- Views and Downloads:** Statistics showing 0 views and 0 downloads.
- Versions:** A table showing 'Version v1' with a date of 'Aug 9, 2025'.
- Communities:** A section indicating that the record is not included in any communities yet.
- Details:** Resource type (Person), Publisher (observations).
- Citation:** A section showing the citation 'Lehenmeier, C. (2025). 1813 248. observations.' with a dropdown menu for the style (APA).
- Export:** A section with a dropdown menu for the export format (JSON) and an 'Export' button.

1: <https://inveniordm.web.cern.ch/>

Bereitstellung des Workflows

SSH Open Marketplace¹

Plattform zum Teilen und Finden von Tools, Services, Datensätzen und Workflows (DARIAH, CLARIN und CESSD)

The screenshot displays the SSH Open Marketplace interface for a workflow titled "Layout recognition of historical tabular records". The workflow is associated with the OCR-D framework. The interface includes a "Details" sidebar on the right with sections for "CATEGORISATION" (Activity: Optical Character Recognition, Keyword: Layout analysis, Discipline: Digital humanities, Language: English, Standard: PAGE XML, Resource category: Digitisation), "Wikidata", and "INFORMATION CONTRIBUTORS" (Constantin Lehenmeier). The main content area shows the "Workflow steps (1)" section, which includes a step titled "1 Detection of region elements" with a "Collapse" button. The description for this step states: "In this step layout regions can be detected using the YOLO extension of OCR-D. A certain YOLO model needs to be provided as well as the mapping from YOLO classes to PAGE XML elements."

1: <https://marketplace.sshopencloud.eu>

Bewertung und Anschlussfähigkeit der Wetterdaten

Kontextualisierung

- Dokumentation historischer Messpraktiken und instrumentaler Fehl-Kalibrierungen (vgl. Winkler, 2020, S. 161)

Validierung

- Prüfung auf Kohärenz und Identifikation auffälliger Ausreißer

Integration

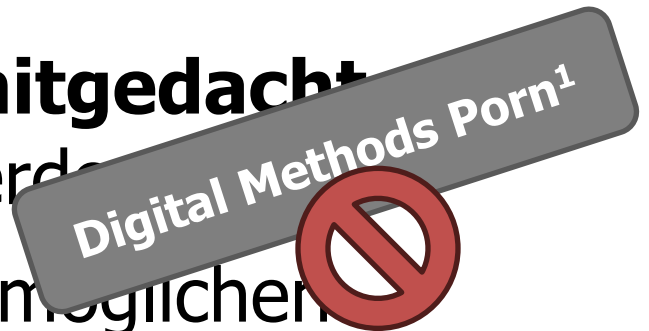
- Einbindung der Daten im *HCLIM*-Datenbestand („The global historical climate database HCLIM“, Lundstad et al., 2023)

Fazit

- 1. Nachnutzbarkeit muss von Anfang an mitgedacht** und in allen Teilschritten des Projekts mitgeplant werden.
- 2. Standardformate statt Insellösungen** ermöglichen interdisziplinäre Vernetzung und Nachnutzung von Forschungsdaten. (*PageXML, TEI*)

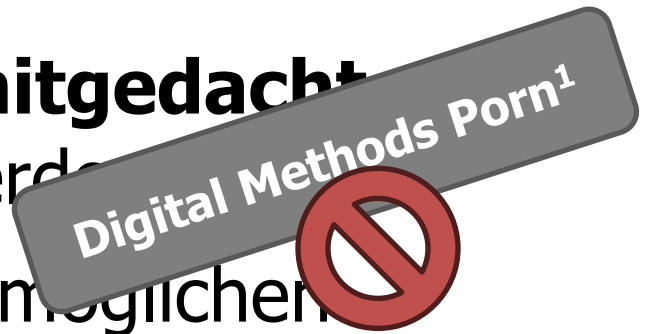
Fazit

- 1. Nachnutzbarkeit muss von Anfang an mitgedacht**
allen Teilschritten des Projekts mitgeplant werden
- 2. Standardformate statt Insellösungen** ermöglichen
interdisziplinäre Vernetzung und Nachnutzung von
Forschungsdaten. (*PageXML, TEI*)



Fazit

- 1. Nachnutzbarkeit muss von Anfang an mitgedacht**
allen Teilschritten des Projekts mitgeplant werden.
- 2. Standardformate statt Insellösungen** ermöglichen
interdisziplinäre Vernetzung und Nachnutzung von
Forschungsdaten. (*PageXML, TEI*)
- 3. Quellenkritik** an historischen Daten bleibt zentral, muss aber um
Modell- und Datenkritik erweitert werden.



**VIELEN DANK FÜR IHRE
AUFMERKSAMKEIT**



<https://github.com/CrazyCrud>



Literaturverzeichnis

Baillot, A. (2024). Digital Methods Porn, oder: Warum die Welt ohne nachnutzbare Workflows nicht zu retten ist. <https://www.germanistik.uni-rostock.de/forschung/digital-humanities/rosdh/ringvorlesung/2024/n/digital-methods-porn-oder-warum-die-welt-ohne-nachnutzbare-workflows-nicht-zu-retten-ist-188645/>. Abgerufen am 22.02.2026

Blömer, J. (2022). *Empirie im Mönchsgewand*. Vandenhoeck & Ruprecht

Clérice, T. (2023). You Actually Look Twice At it (YALTAi): using an object detection approach instead of region segmentation within the Kraken engine. *Journal of Data Mining & Digital Humanities, Historical Documents and...*

Heinrich, P. (1789). Abhandlung über die Wirkung des Geschützes auf Gewitterwolken, welche 1788 den Preis erhalten hat

Kwon, H., An, J., Lee, D., & Shin, W.-Y. (2022). DATa: Domain Adaptation-aided deep Table detection using visual–lexical representations. *Knowledge-Based Systems*, 258

Lundstad, E., Brugnara, Y., Pappert, D., Kopp, J., Samakinwa, E., Hürzeler, A., ... Brönnimann, S. (2023). The global historical climate database HCLIM. *Scientific Data*

Mauelshagen, F. (2016). Ein neues Klima im 18. Jahrhundert. *Zeitschrift für Kulturwissenschaften*, 10 (1), 39-58

Pollin, C. (2023). Wie hältst du's mit der historischen Information in digitalen Editionen?" Workflows, Semantic Web und TEI in GAMS.

https://docs.google.com/presentation/d/1hiyZ7qZgdZEeEUusU_OadQrNyhmLnSt2JBxLRjQKSRA/edit?slide=id.g2485c1e3c1d_0_69#slide=id.g2485c1e3c1d_0_69. Abgerufen am 22.02.2026

Weber, D. (2024, September 12). On the Historiographic Authority of Machine Learning Systems. Digital History Switzerland 2024 (DigiHistCH24), University of Basel.

Winkler, P. (2020). Geschichte der Meteorologie in Bayern bis zum Jahr 1900. In *Geschichte der Meteorologie in Deutschland*

Vogeler, G. (2017). Digitale Edition: Grundlagen und Workflow. https://www.i-d-e.de/wp-content/uploads/2017/11/01_Digitale_Editionen_Vogeler.pdf. Abgerufen am 22.02.2026