



Kollaboratives Zeitunglesen und -annotieren: Das Wien(n)erische Diarium als Fallbeispiel

Claudia Resch, Daniel Schopper, Dario Kampkaspar, Günter Mühlberger

Wien(n)erisches Diarium 1703 - 1779, seit 1780 Wiener Zeitung



Für meine Disziplin ist das Wien[n]erische Diarium eine wertvolle Quelle, weil

... es wertvolle **Personendaten** und Kontextinformationen zu den von uns im Projekt erhobenen (Wiener) Ehefrauen und Ehemännern beinhaltet.

... wir aus den zeitnahen Berichten Details über **Auftraggeber**, **Entscheidungsstände**, **Beteiligte** und zeitgenössische Wahrnehmung erhalten.

... es Hintergründe über die **Medien- und Sozialgeschichte** im Kontext der **Literatur** bietet und dabei weitgehend die obere Seite des sozialen Machtgefälles wiedergibt.

... ich **linguistisch** forschen möchte.

... ANNO eine einfache Möglichkeit darstellt, zeitgenössische **Rezensionen** zu finden.

... über **Tagesabläufe der kaiserl. Familie** und **Ereignisse** berichtet wird.

... es u. a. Informationen über die Verwendung von (Preis- und Grundstein-) **Medaillen** liefert, von der Einführung neuer **Münzen** bzw. Entwertung ausländischer Münzen berichtet.

... **Druckschriften** und **Kupferstiche** angekündigt und beworben werden.

... oft singuläre Nachrichten über **Musikleben am Wiener Hof** zu finden sind.

... in der **Familienforschung** die zeitgenössische Informationslage sehr hilfreich ist.

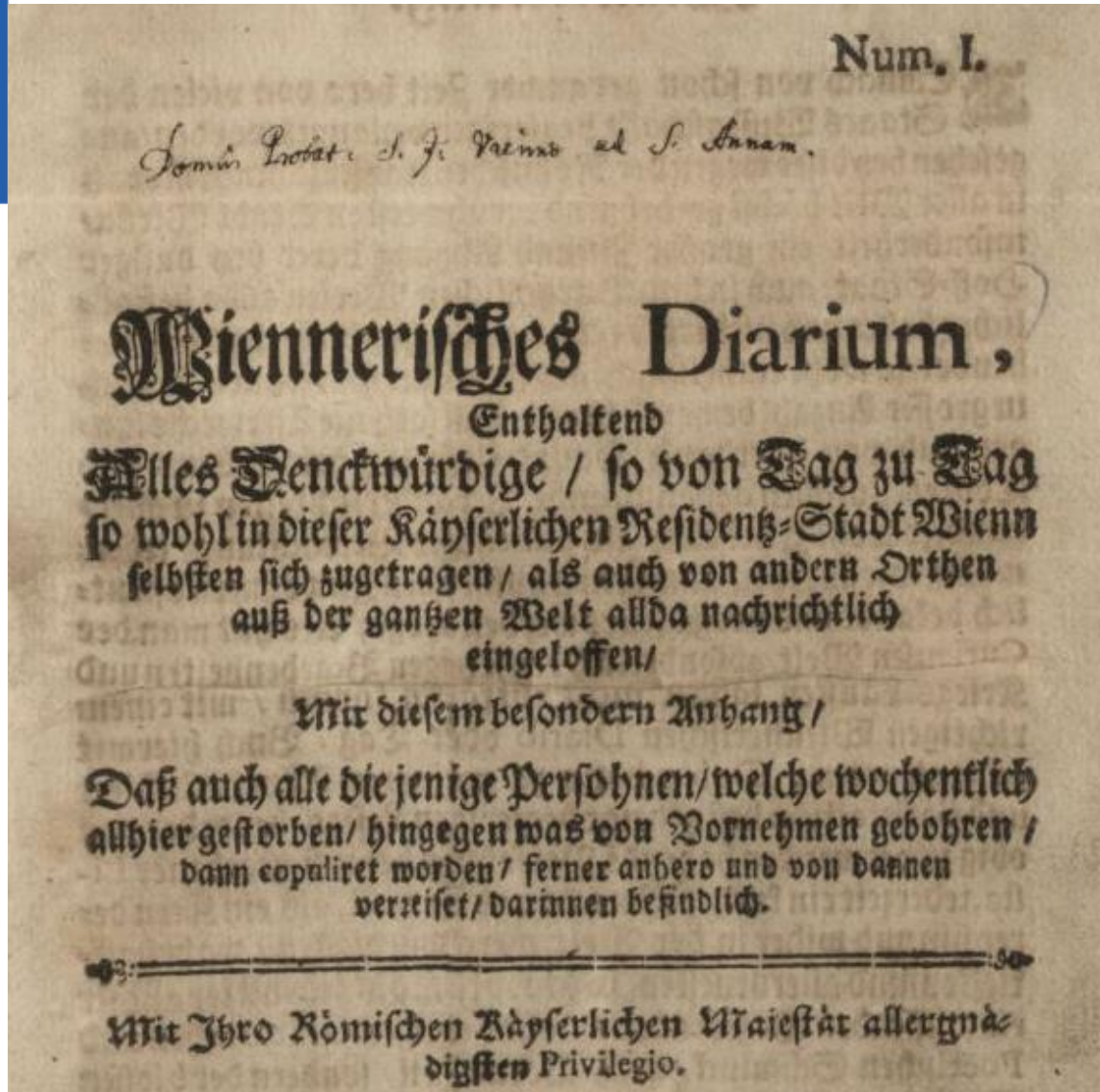
... darin Informationen zu den **Hinrichtungen** in Wien und den Beteiligten enthalten sind.

ANNO - AustriaN Newspapers Online

Historische österreichische Zeitungen und Zeitschriften online

➔ Suchen

→ **neu** ANNO-Suche: Volltextsuche in historischen Zeitungen (1689-1946)



Num. I.

lotb* - -/, B\ fv-ntSt <s >^»^» .

icnrm

[eö Diarium,

Enthaltend

Uenckwürdige / so von Mag zu Wag

so wohlmdteser Kayserlichen Residentz-Stadt Wienn

stlbsten sich zugetragen / als auch von andern Orthen

auß der gantzen Welt allda nachrichtlich

eingeloffen,

Mir dieftmbesonder» Anhantt,

Daß auch ave die jenkgtpersohnen/welche wöchentlich

«llhier gestorben/ hingegen was von Vornehmen gebohren /

dann copuliret worden /ferner anbero und von dannen

verreiset/darinnen befindlich.

iffi«-' ■=. r- ' \.-`Str-teto

Mir Ihrs Römischen Leserlichen ttsajestür ttllergrrL-

digsten?nvLe§jo.



Projektdaten

Förderung: GO!DIGITAL Call 2.0

Titel: **Das Wien(n)erische Diarium:
Digitaler Datenschatz für die geisteswissenschaftlichen Disziplinen**

Laufzeit: 1. März 2017-28. Februar 2019 (= 2 Jahre)

Projektpartner:

- ◆ Austrian Centre for Digital Humanities - Projektleitung (ÖAW)
- ◆ Institut für kunst- und musikhistorische Forschungen (ÖAW)
- ◆ Digitalisierung & Elektronische Archivierung
(Universität Innsbruck / Forschungszentrum Digital Humanities)
- ◆ Innsbruck University Innovations



Projektziele

- 1 mehrere hundert Ausgaben des Diariums (18. Jahrhundert) im Volltext zur Verfügung zu stellen
 - 2 mit geisteswissenschaftlichen Fachdisziplinen* beraten, mit welchen Informationen der Volltext angereichert werden soll
- * Historische Medienforschung, Geschichte, Kunstgeschichte, Medizingeschichte, Sprachgeschichte, Literaturwissenschaft, Musikgeschichte, Zeremonialforschung, Namenforschung, etc.
- 3 Entwicklung einer digitalen Annotations- und Publikationsplattform („Digitalarium“)

Workflow



1. Digitalisierte Bilder beschaffen
2. Texterkennung
3. Einstellen in Online-Präsentation

Workflow



1. Digitalisierte Bilder beschaffen

- Digitalisate der einzelnen Ausgaben
- Ein Bild je Seite
- Gute Qualität, aber nicht zu groß

Workflow



1. Digitalisierte Bilder beschaffen

1. Herunterladen aus ANNO (automatisch)

- Liste mit allen benötigten Ausgaben erstellen
- URL eintragen

- ANNO liefert eine PDF je Ausgabe, daher Einzelbilder extrahieren

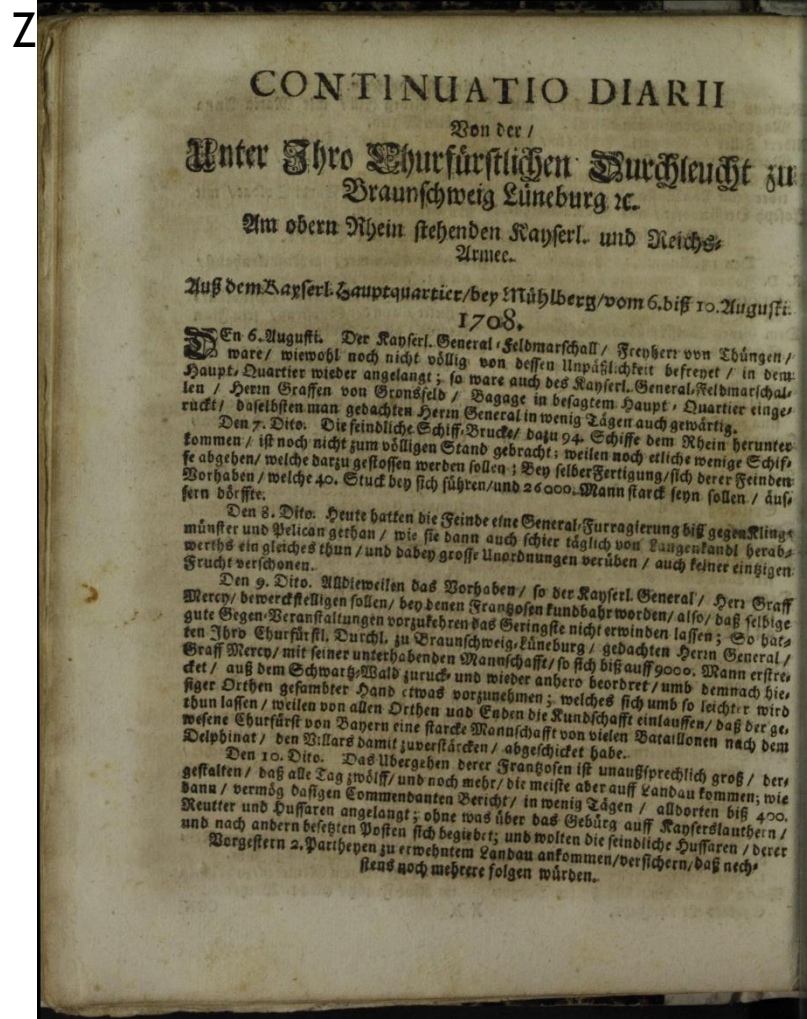
Workflow



1. Digitalisierte Bilder beschaffen

1. Herunterladen aus ANNO (automatisch)
2. Bildkorrekturen (automatisch)

- Bilder sind ggf. schief; die automatische Erkennung benötigt jedoch möglichst horizontale



Workflow



1. Digitalisierte Bilder beschaffen
 1. Herunterladen aus ANNO (automatisch)
 2. Bildkorrekturen (automatisch)
 3. Hochladen in die Software (automatisch)
 2. Texterkennung
 1. Erstellen der Grundlage (automatisch)
- Upload der bearbeiteten Bilddateien per FTP, jeweils ein Verzeichnis je Ausgabe
 - Software: Transkribus
 - Erstellen eines Dokuments aus den Einzelbildern, setzen der Metadaten

Workflow



1. Digitalisierte Bilder beschaffen
 1. Herunterladen aus ANNO (automatisch)
 2. Bildkorrekturen (automatisch)
 3. Hochladen in die Software (automatisch)
2. Texterkennung
 1. Erstellen der Grundlage (automatisch)
 2. Layout-Erkennung (automatisch)
 3. Schrifterkennung (automatisch)

- Erkennen von Graphiken, Trennlinien, Textregionen, Zeilen, Grundlinien
- Grundlinien dienen als Grundlage der Erkennung des Textes und müssen möglichst genau sein
- Anstatt »klassischer« OCR (optical character recognition) wird hier HTR (handwritten text recognition) verwendet, um die Erkennung von Fraktur zu verbessern

Workflow



1. Digitalisierte Bilder beschaffen
 1. Herunterladen aus ANNO (automatisch)
 2. Bildkorrekturen (automatisch)
 3. Hochladen in die Software (automatisch)
2. Texterkennung
 1. Erstellen der Grundlage (automatisch)
 2. Layout-Erkennung (automatisch)
 3. Schrifterkennung (automatisch)
 4. *Korrekturen an einzelnen Bildern*
 5. *Schrifterkennung an fehlgeschlagenen Seiten*

Notwendige Korrekturen sind u.a.

- Weiteres Geraderichten
- Entzerren
- Beschneiden

- Wiederholung von 2.2 und 2.3 auf den betroffenen Seiten
- Ggf. mehrere Durchläufe notwendig

Workflow



1. Digitalisierte Bilder beschaffen
 1. Herunterladen aus ANNO (automatisch)
 2. Bildkorrekturen (automatisch)
 3. Hochladen in die Software (automatisch)
 2. Texterkennung
 1. Erstellen der Grundlage (automatisch)
 2. Layout-Erkennung (automatisch)
 3. Schrifterkennung (automatisch)
 4. *Korrekturen an einzelnen Bildern*
 5. *Schrifterkennung an fehlgeschlagenen Seiten*
 6. *Korrektur durch Dienstleister*
 7. *Kontrolle*
- Texterkennung erreicht Genauigkeiten zwischen 80% und 99%
 - 99% heißt, das etwa eines von 100 Zeichen falsch erkannt ist
 - Bei ca. 3.000 Zeichen je Seite sind also 30 Zeichen falsch erkannt
 - Ziel: mindestens 99,7% Genauigkeit, also ca. 10 falsche Zeichen je Seite

 - Layout-Kennungen richtig vergeben?
 - Stichproben für Textgenauigkeit
 - Teile übersehen?

Workflow



1. Digitalisierte Bilder beschaffen
 1. Herunterladen aus ANNO (automatisch)
 2. Bildkorrekturen (automatisch)
 3. Hochladen in die Software (automatisch)
 2. Texterkennung
 1. Erstellen der Grundlage (automatisch)
 2. Layout-Erkennung (automatisch)
 3. Schrifterkennung (automatisch)
 4. *Korrekturen an einzelnen Bildern*
 5. *Schrifterkennung an fehlgeschlagenen Seiten*
 6. *Korrektur durch Dienstleister*
 7. *Kontrolle*
 3. Einstellen in Online-Präsentation
 1. Export aus Software (automatisch)
 2. Nachbearbeitung (automatisch)
 3. Upload in Web-App (automatisch)
 4. *Kontrolle von Layout etc.*
- Export in ein gewünschtes Zielformat, hier: TEI
 - Kodierung des Layouts
 - Whitespace tokenization

Planung 2018



- Laufende Volltextdigitalisierung
- Arbeit an der Plattform
- Vernetzung:
 - mit anderen Zeitungsdigitalisierungsprojekten
 - in der TEI-SIG „Periodicals and Newspapers“
 - Austausch mit ExpertInnen für das Diarium:
 - Annotate-a-thons
 - Konferenz: „Vom Diarium zum Digitalium“ 24./25.04.2018

734 fertig, 1646 in Arbeit von 7500 Seiten

38 fertig, 98 in Arbeit von 423 Ausgaben



<https://www.oeaw.ac.at/acdh/projects/wienerisches-diarium-digital/>

Literatur (in Auswahl):

Wien(n)erisches Diarium / Wiener Zeitung 1703-1799. In: ANNO - Austrian Newspapers Online.
<<http://anno.onb.ac.at/cgi-content/anno?aid=wrz>> [1. Dezember 2017]

Berger, Martha: „Wienerisches Diarium“ 1703-1780: ein Beitrag zur Entwicklung des Verhältnisses zwischen Staat und Presse. Wien, phil. Diss. 1995.

Blühm, Elger: Zeitungswesen im 17. Jahrhundert. Internationaler Arbeitskreis für Barockliteratur 1.
Wolfenbüttel 1973.

Duchkowitsch, Wolfgang: Zeitung und Bibliothek. Der Stand der Erschließung österreichischer Zeitungen des 17. und 18. Jahrhunderts und Vorstellungen für den Soll-Zustand. In: Das historische und wertvolle Buch in der Bibliotheksverwaltung. Wien 1980.

Gooding, Paul: Historic Newspapers in the Digital Age. “Search All About It!” London/New York: Routledge 2017.

Lang, Helmut W.: Die Anfänge der periodischen Presse in Österreich. In: Bibliographie der österreichischen Zeitungen 1621-1945, Band 2. München: K. G. Saur 2003. S. 21-29.

Reisner, Andrea und Alfred Schiemer: Das Wien(n)erische Diarium und die Entstehung der periodischen Presse. In: Österreichische Mediengeschichte. Wiesbaden: Springer 2016. S. 87-112.

Stamprech, Franz: Die älteste Tageszeitung der Welt: Werden und Entwicklung der „Wiener Zeitung“. Wien: Verlag der Österreichischen Staatsdruckerei 1977.

Schuster, Britt-Marie und Manuel Wille: Von der Kanzlei- zur Bürgersprache? Textsortengeschichtliche Betrachtungen zur „Staats- und gelehrten Zeitung des Hamburgischen unpartheyischen Correspondenten im 18. Jahrhundert“. In: Jahrbuch für Kommunikationsgeschichte 17. Stuttgart: Franz Steiner Verlag 2016. S. 7-29.

Wilke, Jürgen: Inhalt und Form der Zeitung im Wandel. In: 400 Jahre Zeitung. Die Entwicklung der Tagespresse im internationalen Kontext. Bremen: Edition Lumière 2008. S. 355-378.

