ÖAW ÖSTERREICHISCHE AKADEMIE DER WISSENSCHAFTEN

austrian centre for digital humanities & cultural heritage

The DIGITARIUM as a research corpus: New approaches to extracting and linking named entities from historical newspapers

Nina C. Rastinger, Matthias Schlögl, and Claudia Resch

Austrian Centre for Digital Humanities and Cultural Heritage

Austrian Academy of Sciences

Enrichment of historical newspapers as general goal

Key question: How can digital collections of historical newspapers be further enriched using data from other digital portals, data collections, and knowledge resources?

Concrete case: Wien[n]erisches DIGITARIUM

- digital full text corpus of 332 issues of the historical „Wiener Zeitung“ from the 18th century
  - 1703-1780: „Wien[n]erisches Diarium“
- ~ 6.000 pages / ~ 3 mil. Tokens
- fulltext creation with Transkribus
- XML format according to TEI-standard
- structural and typographical annotation

→ project: https://digitarium.acdh.oeaw.ac.at

→ corpus: https://digitarium-app.acdh.oeaw.ac.at (Resch & Kampkaspar 2020)

Wienerisches DIGITARIUM.

Nr. 1, [2.–7. August 1703]

Nr. 2, 8.–11. August 1703

Vienna Time Machine

→ PI: Claudia Resch

→ diachronic connections between people, places & events in Vienna

→ *Wien[n]erisches Di(git)arium* as valuable source material about Vienna in the Early Modern Period

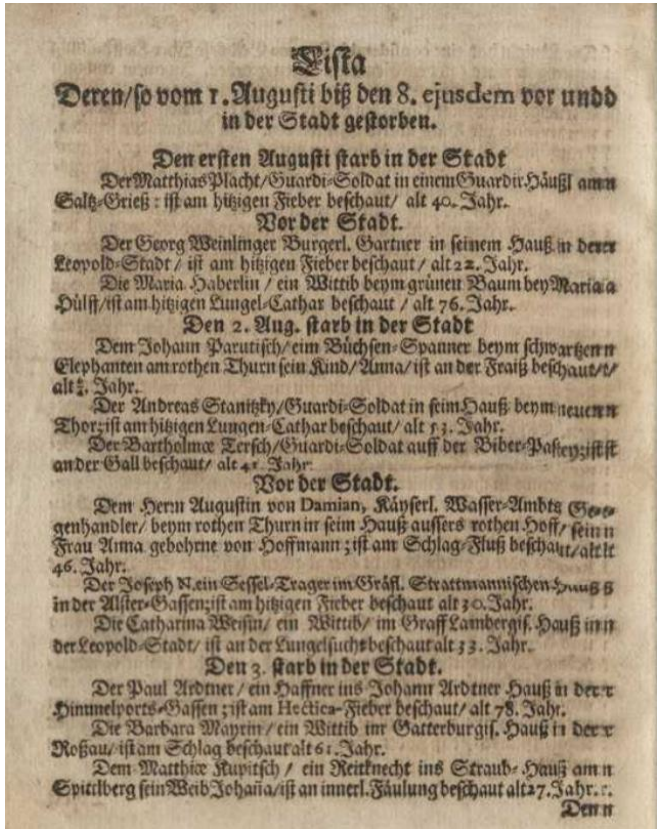
Stadt Wien

Time Machine

Starting point: obituaries

→ lists of deceased persons (inside and outside of the city centre of Vienna)

- published in almost all issues of the historical *Wiener Zeitung*
- relatively consistent structure
- contain detailed information on name, occupation, place of death and age
- occasionally also title, marital status and cause of death given



Liſta
Deren/ſo vom 1. Auguſti biß den 8. ejusdem vor undd in der Stadt geſtorben.

Den erſten Auguſti ſtarb in der Stadt
Der Matthias Placht/Guardi-Soldat in einem Guardir Häußl amn Salz-Grieß: iſt am hitzigen Fieber beſchaut/ alt 40. Jahr.

Vor der Stadt.
Der Georg Weinlinger Burgerl. Gartner in ſeinem Hauß in derr Leopold-Stadt/ iſt am hitzigen Fieber beſchaut/ alt 22. Jahr.
Die Maria Haberlin/ ein Wittib beym grünen Baum bey Maria a Hülf/iſt am hitzigen Lungel-Cathar beſchaut/ alt 76. Jahr.

Den 2. Aug. ſtarb in der Stadt
Dem Johann Parutiſch/eim Büchſen-Spanner beym ſchwartzenn Elephanten am rothen Thurn ſein Kind/Anna/iſt an der Fraiß beſchaut/ alt 3. Jahr.
Der Andreas Stanitzky/Guardi-Soldat in ſeim Hauß beym neuenn Thor: iſt am hitzigen Lungen-Cathar beſchaut/ alt 53. Jahr.
Der Bartholmæ Terſch/Guardi-Soldat auff der Biber-Paſtey iſt an der Gall beſchaut/ alt 41. Jahr.

Vor der Stadt.
Dem Herrn Auguſtin von Damian, Käyſerl. Waſſer-Ambts Gegenhandler/ beym rothen Thurn in ſein Hauß auſſers rothen Hoff/ ſeinn Frau Anna gebohrne von Hoffmann: iſt am Schlag-Fluß beſchaut/ alt 46. Jahr.
Der Joſeph N. ein Seſſel-Trager im Gräfl. Strattmanniſchen Hauß in der Alſter-Gaſſen: iſt am hitzigen Fieber beſchaut alt 30. Jahr.
Die Catharina Weiſſin/ ein Wittib/ im Graff Lambergiſ. Hauß in der Leopold-Stadt/ iſt an der Lungelſucht beſchaut alt 33. Jahr.

Den 3. ſtarb in der Stadt.
Der Paul Ardtner/ ein Haffner ins Johann Ardtner Hauß in derr Himmelports-Gaſſen: iſt am Hectica-Fieber beſchaut/ alt 78. Jahr.
Die Barbara Mayrin/ ein Wittib im Gatterburgiſ. Hauß in derr Roßau/ iſt am Schlag beſchaut alt 61. Jahr.
Dem Matthiæ Kupitſch/ ein Reitknecht ins Straub-Hauß amn Spitelberg ſein Weib Johanna/iſt an innerl. Fäulung beſchaut alt 27. Jahr.

Dem

Starting point: obituaries

Wienerisches DIGITARIUM.

Wiennerisches Diarium

Nr. 1, [2.–7. August 1703]

[Navigation einblenden] [Suche] [Hilfe]

‖ ← [8] → ‖

Lista
Deren / so vom 1. *Augusti* biß den 8. *ejusdem* vor und in der Stadt gestorben.

Den ersten Augusti starb in der Stadt

Der Matthias Placht / Guardi=Soldat in einem Guardir Häußl am Saltz=Grieß: ist am hitzigen Fieber beschaut / alt 40. Jahr.

Vor der Stadt.

Der Georg Weinlinger Burgerl. Gartner in seinem Hauß in der Leopold=Stadt / ist am hitzigen Fieber beschaut / alt 22. Jahr.

Die Maria Haberlin / ein Wittib beym grünen Baum bey Maria Hülff / ist am hitzigen Lungel=Cathar beschaut / alt 76. Jahr.

Den 2. Aug. starb in der Stadt

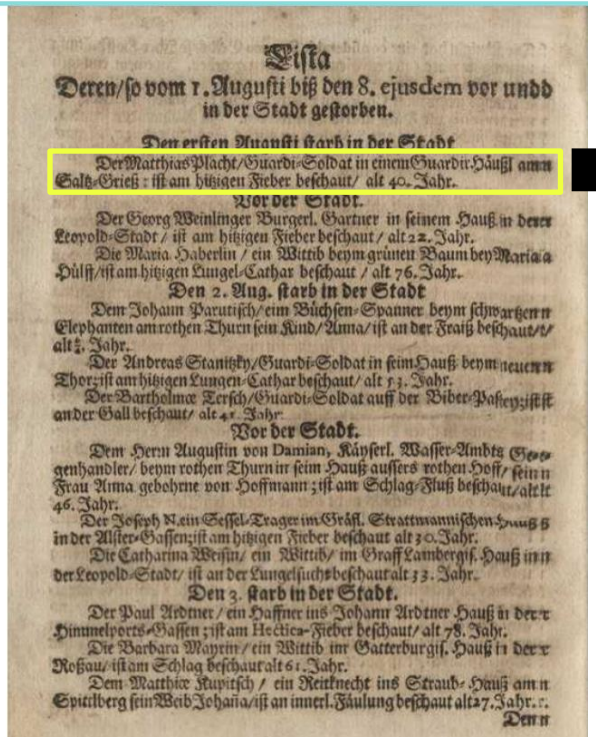
Dem Johann Parutisch / eim Büchsen=Spanner beym schwartzen Elephanten am rothen Thurn sein Kind / Anna / ist an der Fraiß beschaut /
alt 6/4. Jahr.

Der Andreas Stanitzky / Guardi=Soldat in seim Hauß beym neuen Thor ist am hitzigen Lungen=Cathar beschaut / alt 53. Jahr.

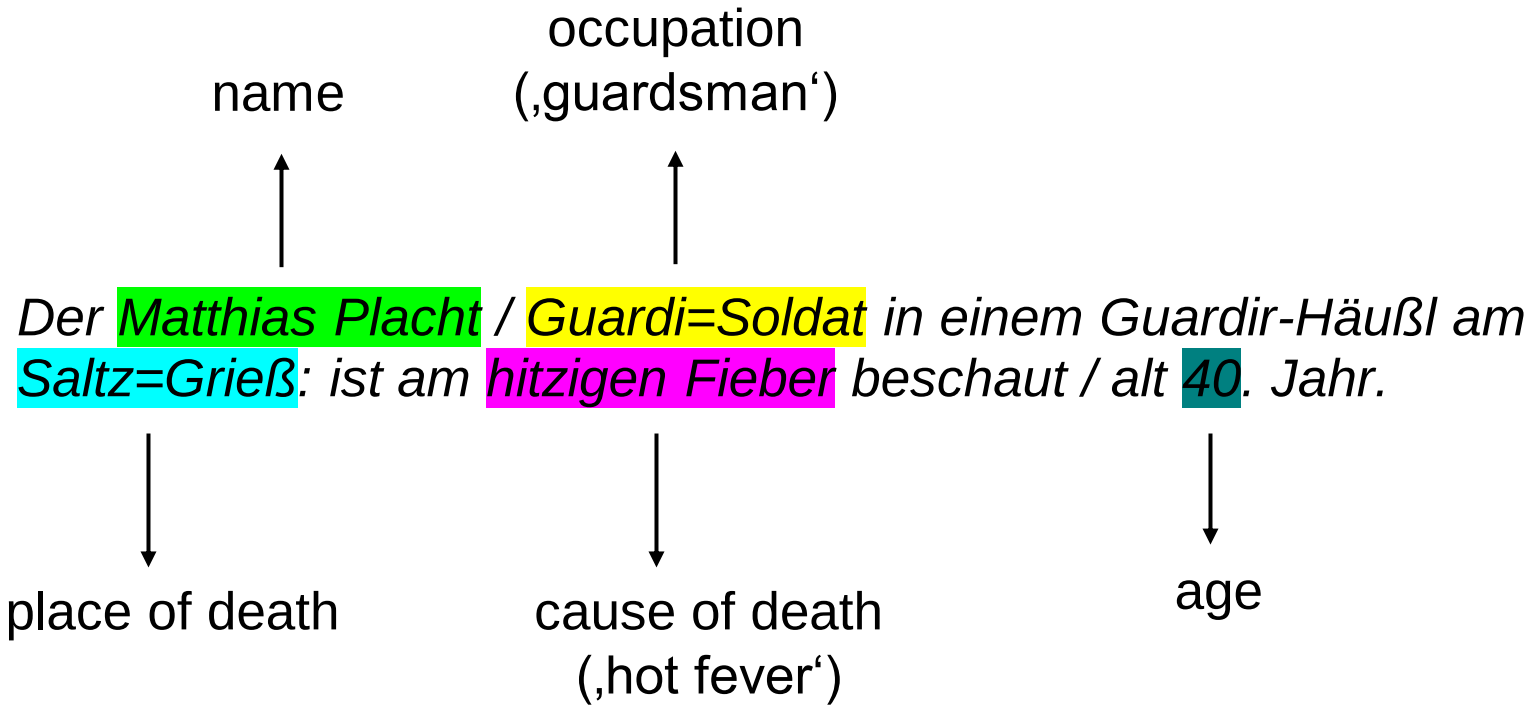
Der Bartholmæ Tersch / Guardi=Soldat auff der Biber=Pastey; ist an der Gall beschaut / alt 41. Jahr.

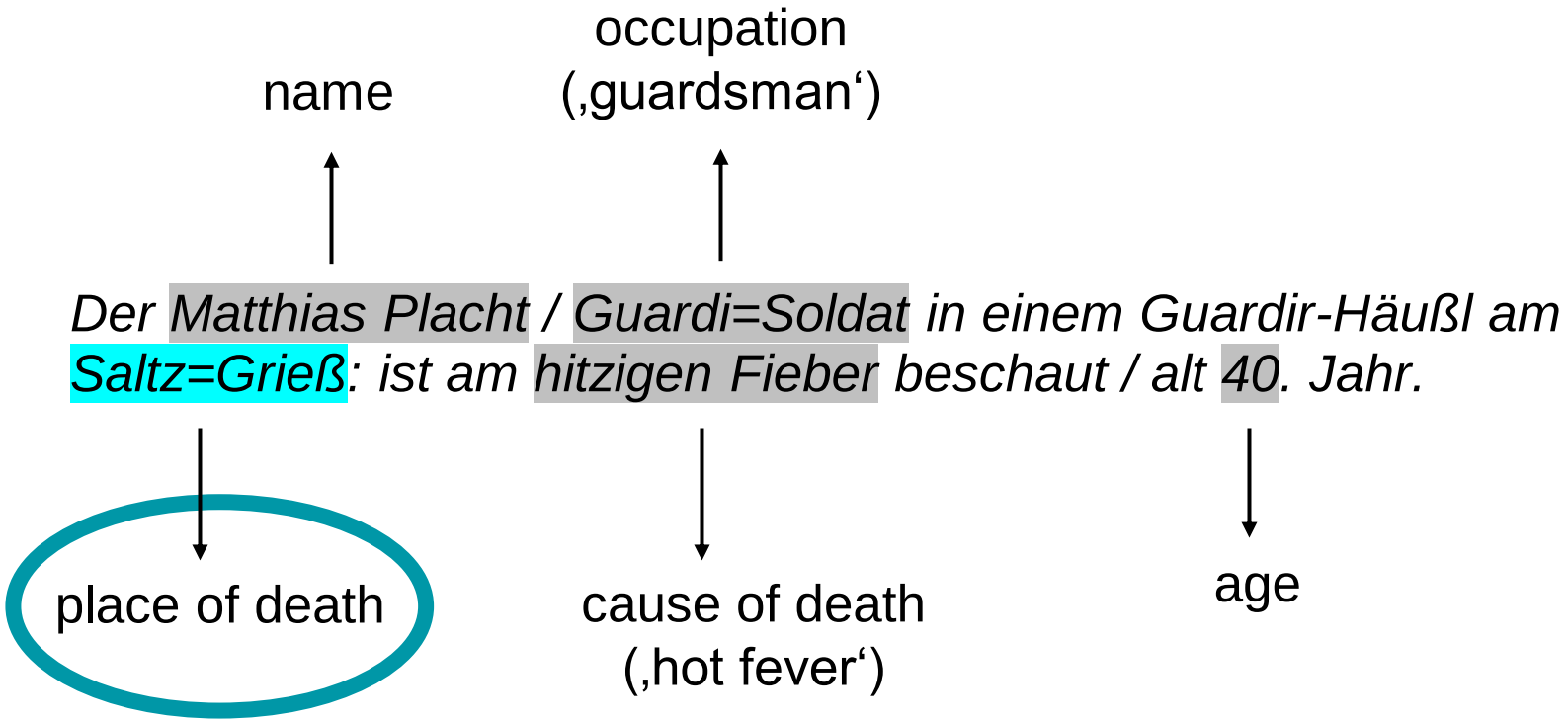
Vor der Stadt.

Dem Hern Augustin von *Damian*, Käyserl. Wasser=Ambts Ge= genhandler / beym rothen Thurn in seim Hauß aussers rothen Hoff / sein


Lista
Deren/ſo vom 1. Auguſti biß den 8. ejusdem vor undd in der Stadt geſtorben.
Den erſten Auguſti ſtarb in der Stadt
Der Matthias Placht/Guardi-Soldat in einem Guardir-Häußl amm Salts-Grieß: iſt am hitzigen Fieber beſchaut/ alt 40. Jahr.
Vor der Stadt.
Der Georg Weinlinger Burgerl. Gartner in ſeinem Hauß in derer Leopold-Stadt/ iſt am hitzigen Fieber beſchaut/ alt 22. Jahr.
Die Maria Haberlin/ ein Wittib beym grünen Baum bey Maria Hülff/ iſt am hitzigen Lungel-Cathar beſchaut/ alt 76. Jahr.
Den 2. Aug. ſtarb in der Stadt
Dem Johann Parutiſch/ eim Büchſen-Spanner beym ſchwartzen Elephanten am rothen Thurn ſein Kind/ Anna/ iſt an der Fraiß beſchaut/ alt 6/4. Jahr.
Der Andreas Stanitzky/ Guardi-Soldat in ſeim Hauß beym neuen Thor; iſt am hitzigen Lungen-Cathar beſchaut/ alt 53. Jahr.
Der Bartholmæ Terſch/ Guardi-Soldat auff der Biber-Paſtey; iſt an der Gall beſchaut/ alt 41. Jahr.
Vor der Stadt.
Dem Hern Auguſtin von Damian, Käyſerl. Waſſer-Ambts Gegenhandler/ beym rothen Thurn in ſein Hauß auſſers rothen Hoff/ ſein Frau Anna gebohrne von Hoffmann; iſt am Schlag-Fluß beſchaut/ alt 46. Jahr.
Der Joſeph N. ein Seſſel-Trager im Gräfl. Strattmanniſchen Hauß in der Alſter-Gaſſen; iſt am hitzigen Fieber beſchaut alt 30. Jahr.
Die Catharina Weiſin/ ein Wittib/ im Graff Lambergiſ. Hauß in der Leopold-Stadt/ iſt an der Lungelſuchts beſchaut alt 33. Jahr.
Den 3. ſtarb in der Stadt.
Der Paul Ardtner/ ein Haffner ins Johann Ardtner Hauß in derer Himmelpforts-Gaſſen; iſt am Hectica-Fieber beſchaut/ alt 78. Jahr.
Die Barbara Mayrin/ ein Wittib im Gatterburgiſ. Hauß in der Roßau/ iſt am Schlag beſchaut alt 61. Jahr.
Dem Matthiæ Kupitſch/ ein Reitknecht ins Strauh-Hauß am Spitelberg ſein Weib Johana/ iſt an innerl. Fäulung beſchaut alt 27. Jahr.
Den

```
<item facs="#facs_8_r_1_3" rendition="#f"
rend="findent" xml:id="i5">
    <lb facs="#facs_8_r1003" xml:id="z241"/>
    <w xml:id="w2551">Der</w>
    <w xml:id="w2552">Matthias</w>
    <w xml:id="w2553">Placht</w>
    <pc xml:id="w2554">/</pc>
    <w xml:id="w2555">Guardi</w>
    <pc xml:id="w2556">=</pc>
    <w xml:id="w2557">Soldat</w>
    <w xml:id="w2558">in</w>
    <w xml:id="w2559">einem</w>
    <w xml:id="w2560">Guardir</w>
    <w xml:id="w2561">Häußl</w>
    <w xml:id="w2562">am</w>
    <lb facs="#facs_8_r1004" xml:id="z242"/>
    <w xml:id="w2563">Saltz</w>
    <pc xml:id="w2564">=</pc>
    <w xml:id="w2565">Grieß</w>
    <pc xml:id="w2566">:</pc>
    <w xml:id="w2567">ist</w>
    <w xml:id="w2568">am</w>
    <w xml:id="w2569">hitzigen</w>
    <w xml:id="w2570">Fieber</w>
    <w xml:id="w2571">beschaut</w>
    <pc xml:id="w2572">/</pc>
    <w xml:id="w2573">alt</w>
    <w xml:id="w2574">40</w>
    <pc xml:id="w2575">.</pc>
    <w xml:id="w2576">Jahr</w>
    <pc xml:id="w2577">.</pc>
</item>
```


name
occupation
(‚guardsman‘)
Der Matthias Placht / Guardi=Soldat in einem Guardir-Häußl am
Saltz=Grieß: ist am hitzigen Fieber beschaut / alt 40. Jahr.
place of death
cause of death
(‚hot fever‘)
age


name
occupation
(‚guardsman‘)
Der Matthias Placht / Guardi=Soldat in einem Guardir-Häußl am Saltz=Grieß: ist am hitzigen Fieber beschaut / alt 40. Jahr.
place of death
cause of death
(‚hot fever‘)
age

1) Named Entity Recognition (NER) Task: systematic extraction of this information from full text

2) Named Entity Linking (NEL) Task: external linking of extracted location data to corresponding data set

WIEN GESCHICHTE WIKI

Wien Geschichte Wiki (‚Vienna History Wiki‘)

- historical knowledge ressource for Vienna
- currently over 47.000 entries as well as maps and further biographical references

e.g. place names:

- > 11.500 entries for topographical objects
- > 7.500 entries for buildings

→ https://www.geschichtewiki.wien.gv.at

DIGITARIUM + Wien Geschichte Wiki

Den ersten Augusti starb in der Stadt

Der Matthias Placht / Guardi=Soldat in einem Guardir Häußl am Saltz=Grieß: ist am hitzigen Fieber beschaut / alt 40. Jahr.

Den 2. Aug. starb in der Stadt

Dem Johann Parutisch / eim Büchsen=Spanner beym schwartzen Elephanten am rothen Thurn sein Kind / Anna / ist an der Fraiß beschaut / alt 6/4. Jahr.

Der Andreas Stanitzky / Guardi=Soldat in seim Hauß beym neuen Thor ist am hitzigen Lungen=Cathar beschaut / alt 53. Jahr.

Der Bartholmæ Tersch / Guardi=Soldat auff der Biber=Pastey; ist an der Gall beschaut / alt 41. Jahr.

Den 3. starb in der Stadt.

Der Paul Ardtner / ein Haffner ins Johann Ardtner Hauß in der Himmelports=Gassen; ist am *Hectica*-Fieber beschaut / alt 78. Jahr.

Die Barbara Mayrin / ein Witib im Gatterburgis. Hauß in der Roßau / ist am Schlag beschaut alt 61. Jahr.

Dem Matthiæ Kupitsch / ein Reitknecht ins Straub=Hauß am Spittlberg sein Weib Johanna / ist an innerl. Fäulung beschaut alt 27. Jahr.

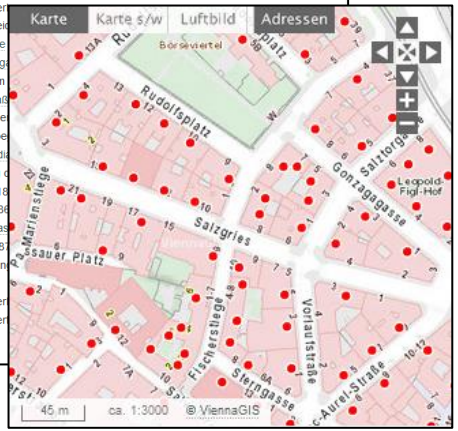
- Salzgries
- Roter Turm
- Neutor
- Biberbastei
- Himmelpfortgasse
- Roßau (Vorstadt)
- Spittelberg (Vorstadt)

Salzgries

Salzgries (1.; Gries). Schon zur Zeit der Römer muss hier eine Uferstraße zwischen Lagermauer und schiffbarem Strom bestanden haben; sie lag noch lange Zeit außerhalb der Stadtmauer.

Die Gegend, ursprünglich "An dem Gries" (sandiges Ufer) genannt, heißt seit 1322 Salzgries. Da hier ("am Gestade") der südlichste Donauarm floss (Donau, Donaukanal), konnten die auf der Donau ankommenden Salzschiffe unmittelbar anlegen. Der daraus resultierende starke Verkehr hatte zur Folge, dass auf dem Salzgries viele Einkehrwirtshäuser entstanden (beispielsweise "Zum Wölfen in der Au", "Zum blauen Hechten", "Zum weißen Löwen") und auch manche Innungshäuser hier verlegt wurden. Die Ringmauer wurde im Bereich Salzgries erst 1661-1664 durch eine Kurtine, welche die Elendbastei und die Große Gonzaga verband. Als im 18. Jahrhundert zwischen dem Rabensteig und der heutigen Marc-Aurel-Straße innerhalb der Kurtine durch den Einschub einer Häuserzeile die Kohlmessergasse entstand, be... Salzgries erst beim Morzinplatz. Beim Concordia... endet er seit 1870 (die nach der Demolierung d... Befestigungsanlagen am Donaukanal [1858-18... erfolgte Einbeziehung der Zeughausgasse [186... bereits 1870 durch die Eröffnung der Börsegas... hinfällig). Nach dem Abbruch des Arsenals (187... und der Salzgrieskaserne (1880), der Eröffnun... Vorlaufstraße (1886) und der Gestaltung des Morzinplatzes (1888) samt ihrer Umgebung ver... Salzgries endgültig sein über Jahrhunderte ver... Gepräge.

Detail Vogelschauplan, Joseph Daniel Huber (1769-1773)



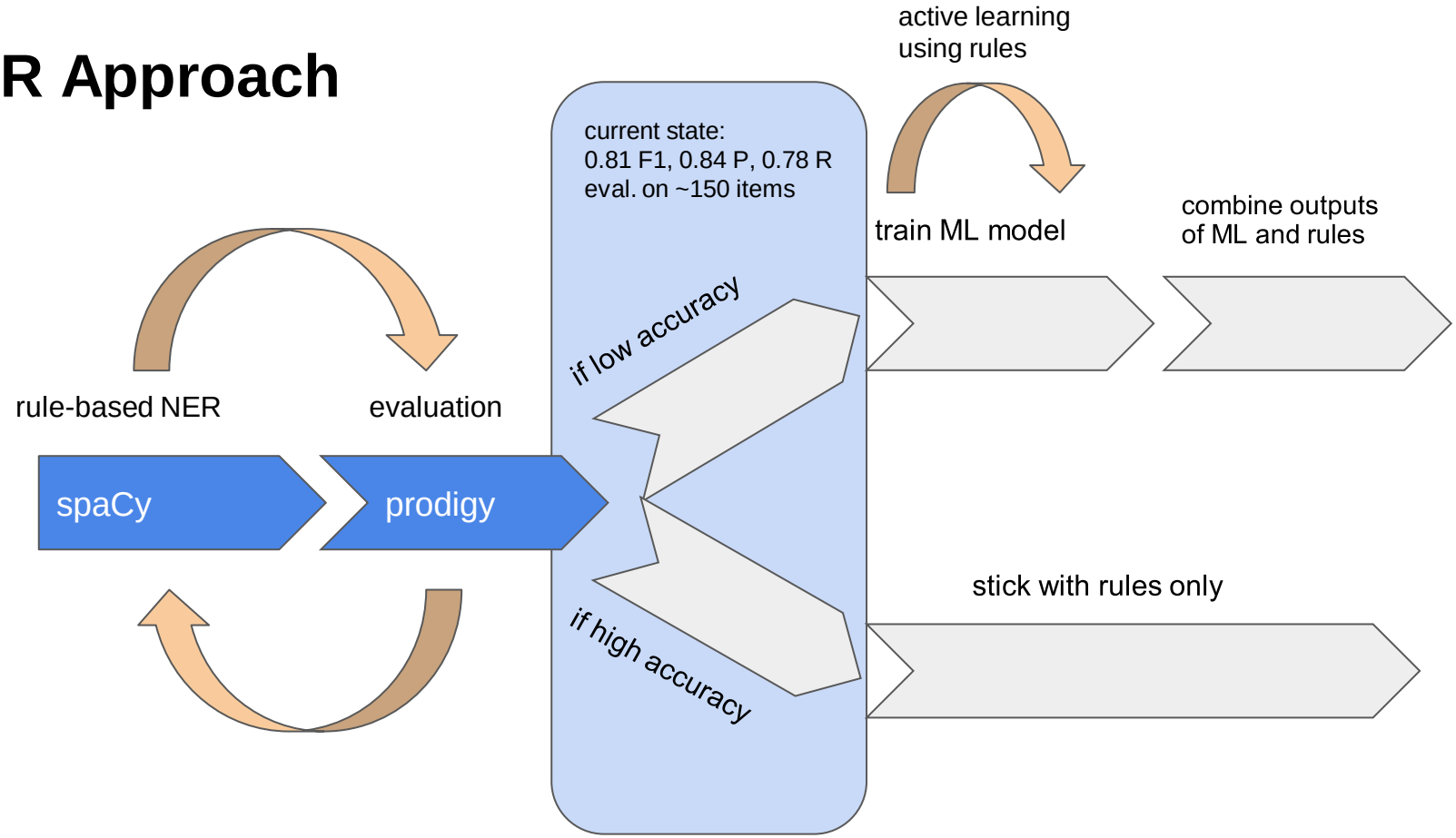
Challenges

- historical variants
  - e. g. *Saltz=Grieß > Salzgries*
  - e. g. *Mariahülff > Mariahilf*
- name changes
  - e. g. *Kärntnerbastei* (1577-1770) > *Augustinerbastei*
- multitude of abbreviations
  - e. g. *i. > in*, *im*, *ihr*, …
  - e. g. *Nicklst. > Nickolsdorf*
  - e. g. *Paar. Gart. > Parisergarten*

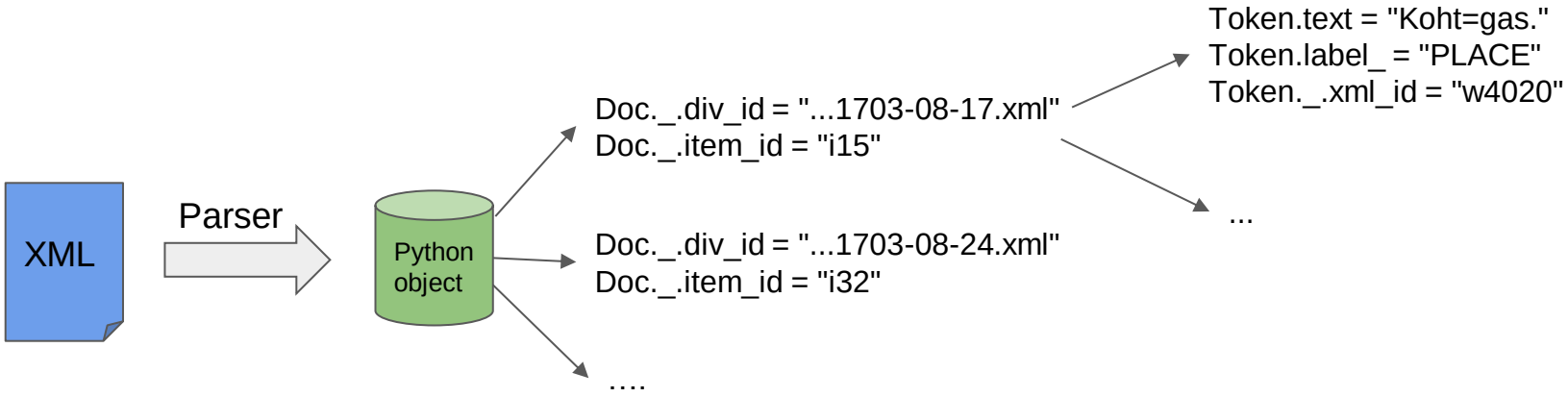
→ problematic for Named Entity Recognizers

DIGITARIUM	WGW
Spitlberg *Spittlberg* *Spitelberg* *Spitalberg* *Spitelb.* …	*Spittelberg*
Cärntner Pastey *Kärntner=pastey* *Cärnther=Pastey* …	*Kärntnerbastei* (*Augustinerbastei*)
Mariahülff *Mariah.* *Mariahülf* *Maria=Hülff* *Mariahilff* …	*Mariahilf*

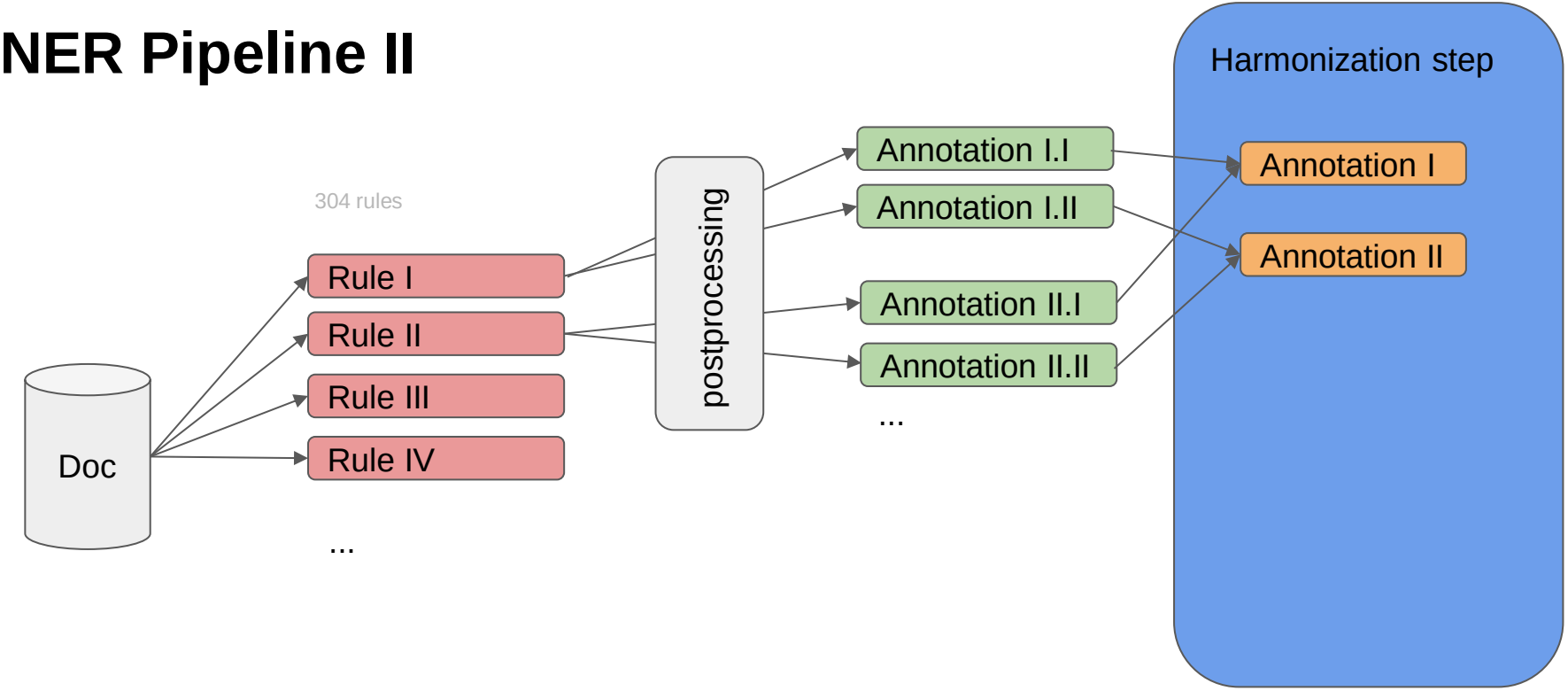
NER Approach



NER Pipeline I

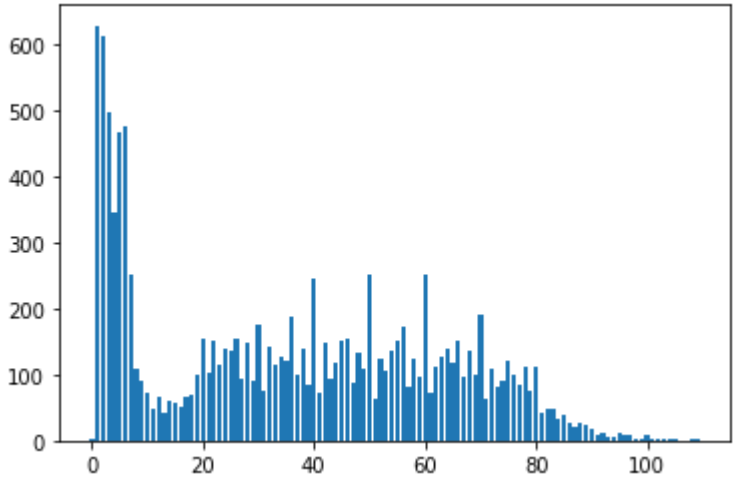


NER Pipeline II



Data

- 13164 entries
- 14877 places annotated
- 6480 unique toponyms
- age for 12120 entries
- between 0 and 109 (!)
- average 35



"Jungfrau Elisabeth Grafensteigerin / in dem Brandweinerischen Haus / bey dem Pauler= Thor / alt 109 . J"

Next steps

- improve rules and XML parser and (re)evaluate
- use rules as bootstrap model for training NER model if F1 < 0.9
- create a named entity linker for places:
  - problem of normalization (*Saltz=grieß => Salzgries*)
  - problem of disambiguation (*Kaisergasse* => 8th district or 15th district)
- create similar pipelines for “cause of death” and “occupation”
- create and publish a LOD dataset

Contact details

Nina C. Rastinger:
ninaclaudia.rastinger@oeaw.ac.at

Matthias Schlögl:
matthias.schloegl@oeaw.ac.at

Claudia Resch:
claudia.resch@oeaw.ac.at