

# Annotation Guidelines for Reichsanzeiger-NLP (NER and NEL tasks)

Renat Shigapov

Lillian Scharnhorst

Alexandra Oldokat

## Contents

Introduction . . . . .	2
General instructions . . . . .	2
Entity types and subtypes . . . . .	2
Components . . . . .	3
Nested entities and special constructions . . . . .	4
Ambiguities . . . . .	4
Entities . . . . .	4
Person (PER) . . . . .	4
Organization (ORG) . . . . .	13
Location (LOC) . . . . .	16
EVENT . . . . .	17
DATE . . . . .	18
PROD.media . . . . .	18
Entity linking . . . . .	18
References . . . . .	19

## Introduction

These guidelines are aimed to support manual annotation of the Reichsanzeiger newspapers for NER & NEL tasks.

**Application context.** Our ultimate goal is to recognize and link the named entities from the historical newspaper Reichsanzeiger. It's roughly 500 thousands pages for years 1819-1945.

The main language of the newspapers is German. The tiny parts are written in French, English, Spanish, Latin and Portuguese.

**Methodology.** These guidelines are created iteratively via: \* Adapting the existing annotation guidelines [1-4] on historical German texts for NER/NEL tasks (mainly [1] based on the annotation guidelines for French texts [5]), \* Analysing the sample pages from the Reichsanzeiger.

## General instructions

Manual annotators shall use only subtypes and components. The only exception is the type EVENT: the events should be annotated as EVENT.

## Entity types and subtypes

The types and their subtypes categorize a named entity. This is the first level of annotation referring to a general segmentation of words into major categories [6]. The taxonomy follows mainly [1] and [5], which are consistent except of the EVENT type (see the Source-column).

Type	Subtypes	The entity refers to	Source
PER	PER.ind	an individual (a proper name should be a part of the entity)	p.21 [5]; p.9 [1]
	PER.coll	more than one individual (a proper name should be a part of the entity)	p.21 [5]; p.9 [1]
ORG	ORG.adm	an organisation which plays a mainly administrative role	p.31 [5]; p.15 [1]
	ORG.ent	an organisation which doesn't play a mainly administrative role	p.29 [5]; p.14 [1]

Type	Subtypes	The entity refers to	Source
LOC	LOC.adm	a territory with a geopolitical border (e.g., cities, city districts, countries & continents)	p.32 [5]; p.15 [1]
	LOC.phys	a physical location (e.g., mountains, rivers & planets)	p.34 [5]; p.17 [1]
	LOC.oro	an oronym (e.g., are streets, squares, roads & highways)	p.35 [5]; p.18 [1]
	LOC.fac	named buildings (train station & museum), named constructions (gates & bridges) & their extensions (stadium & campus); a physical location of an organisation	p.36 [5]; p.19 [1]
	LOC.add	physical & electronic addresses	p.37 [5]; p.20 [1]
PROD	PROD.media	media production (e.g., newspapers, magazines & sales catalogues)	p.41 [5]; p.22 [1]
TIME	TIME.date.abs	an absolute date (specific date, not a relative date; the dates containing only a day and a month, a month and a year, only a year or only a century)	p.57 [5]; p.23 [1]
	TIME.range	a time interval between two absolute dates	-
EVENT	EVENT	an event	p.63 [5]

### Components

Components categorize the elements inside a named entity. This is the second level of annotation helping to determine the named entity type and to set the named entity

boundaries [6]. The components can never be used outside the scope of a type or subtype element [6]. A named entity can consist of one or more components as well as the parts without components. The components of the type PERSON are:

Component	Description
COMP.name	first, middle and last names as well as nickname and initials of a person
COMP.title	title or designator of a person
COMP.func	a function or job of a named person
COMP.qualifier	specifies a person in the form of a qualifying adjective
COMP.demonym	the geographical origin of a person

Annotators shall annotate components only for named entities of type person.

### **Nested entities and special constructions**

**Nested entities.** A nested entity is an entity nested in another entity or in entity component. There are no limits on nesting levels during annotation.

**Components of nested entities.** In contrast to [1], components of nested entities are also annotated.

### **Ambiguities**

**Unsolvable entity type ambiguities.** \* Despite our efforts to annotate ambiguous entity types (e.g., firms named after individuals as both PER.ind and ORG.adm), we found that consistent resolution was not always feasible.

### **Metonymy.**

- Metonymy was explored as a potential linguistic phenomenon in our analysis, but we ultimately found it to be of limited relevance to the specific entities present in our dataset.

## **Entities**

### **Person (PER)**

#### **Subtypes**

- PER.ind: the entity refers to an individual.
- PER.coll: the entity refers to more than one individual. A proper name should be a part of the entity.

## Coverage of the type Person

- Considered as Person:
  - real person (e.g., Karl August von Hardenberg)
  - imaginary characters and characters of literature pieces (e.g., Marthe Schwerdtlein)
  - religious figures (God)
  - titles if they can be distinctively attributed to one person for ex. via date (e.g., Se. Königl. Hoheit der Herzog von Cumberland, Kaiser der Franzosen)
  - firm names that are the name of a person as well
- Not considered as Person: (TO DO: decide on this keeping in mind the entity linking task)
  - expressions without a proper name except expressions containing title and demonym or expressions that can be clearly attributed to one person via time (e.g., Großherzog von Baden, Kaiser der Franzosen)
  - demonyms which do not modify a proper name (e.g., preußischer Staatsbürger)
  - isolated functions not attached to a person name
  - abbreviation of names that are only one letter (e.g., A.)
- Considered as Person.Collective:
  - more than one individual containing a proper name (e.g., )
  - royal courts (e.g., Kaiserlicher Russischer Hofstaat)
  - firms with several partners (e.g.,
- Not considered as Person.Collective:
  - citizens or residents of certain geographic areas (e.g., Herren H.F. Fetschow & Sohn)
  - Löwenberger, Plagwitzer, die letzten Franzosen)
  - families if it is not clear which family members are included
  - members of an institution that frequently change (e.g., Reichstagsabgeordnete)
  - if something is both PER.coll and ORG.ent both entities are added

## specific heuristics for Person

- Quartett + name (music) is annotated PER.coll (e.g., Schwedisches National Damen Doppel Quartett)

## Person Components

- COMP.func (a function or job of a named person):
  - an occupation, profession or specialty (e.g., Zimmermann, Richter)
  - an administrative function in public or private area (e.g., Vorsitzender, Außenminister)
  - titles of advising officials (Wirklicher Geheimer Rat, Landrat, Landrath a. D.)
  - social roles and status (e.g., Häftling)
  - tasks (command, directorate) that describe the function of a specific person (e.g., Führung der 29. Kav. Brig., Höchstkommmandirenden des X. Armee Corps, Direktion, Regie: Emil Hahn )

- roles (for ex. Pastor (both military and civilian), Archimandriten) within the church
- family relations (e.g., Gemahlin / Ehefrau von ..., siblings)
- certain adjectives: ehemalig / früher / emeritiert/ ausgeschiedener/ geschäftsleitend/ verw. (abbreviation for verwitwet = widowed)
- a function always includes the organization, place or specialization attached to it [1]
- COMP.title (title or designator of a person):
  - a civil or honorific prefix (e.g., Frau, Herr, Damen, Herren, Dlle. (demoiselle) Dr., Majestät, königliche Hoheit), military titles (e.g., General, Leutnant), nobility titles and royal titles (Fürstin, Gräfin, Herzog, Ritter, Junker)
  - specifications of doctorates (e.g., Dr. jur., Dr. rer. nat.)
  - titles that a
  - re both civil and military titles (e.g., Kapitän)
  - certain adjectives if a title follows are added to COMP.tilte: ehemalig, früher, emeritiert, ausgeschiedener, Abbreviation "a. D."
  - Functions (Wehrmann, Reservisten, Arbeitssoldat II. Klasse, Einjährig-Freiwilliger) within the military that are not titles are annotated
- COMP.qualifier (specifies a person in the form of a qualifying adjective):
  - any adjective qualifying the entity (e.g., sozialistische, senior, III., verantwortlich, verstorben)
- COMP.name (first, middle and last names as well as nickname and initials of a person):
  - covers first-, middle-, last- and nickname (e.g., Karl)
  - names of noble families if the name is not related to a location (e.g., von Humboldt)
- COMP.demonym (the geographical origin of a person):
  - a noun or adjective that identifies residents of a particular place (e.g., Bayerische)
  - a preposition that identifies the origin of a person (in, aus, zu not in)
  - names of noble families if the name is related to a location (e.g., von Solms-Lych)
  - if locations are separable, they are annotated individually: see example
  - if two locations are not separable, they are annotated together (e.g., RahmRegBez. Düffeldorf) ##### specific heuristics for Person Components
- COMP.func: both civilian and military doctors are annotated as COMP.func

**Tricky cases for Person** Prinzen Karl, Louis, und Ferdinand zu Solms-Lych

```

<PER.coll>
  <COMP.title>Prinzen</COMP.title>
  <PER.ind> Karl</PER.ind>,
  <PER.ind> Louis</PER.ind>
  ,und
  <PER.ind> Ferdinand</PER.ind>
  <COMP.demonym> zu Solms-Lych</COMP.demonym>
</PER.coll>

```

```

<PER.coll>
  <COMP.title> Herren </COMP.title>
  <ORG.ent>
    <PER.ind>H.F. Fetschow </PER.ind>& Sohn
  </ORG.ent>
</PER.coll>

```

The expressions containing title and demonym are annotated:

Großherzog von Baden

```

<PER.ind>
  <COMP.title>Großherzog</COMP.title>
  <COMP.demonym>von Baden</COMP.demonym>
</PER.ind>

```

Se. Königl. Hoheit der Herzog von Cumberland

```

<PER.ind>
  <COMP.title>Se. Königl. Hoheit</COMP.title>
  der
  <COMP.title>Herzog</COMP.title>
  <COMP.demonym>von Cumberland</COMP.demonym>
</PER.ind>

```

Königl. Preußische Lieutenant im Garde-Uhlanen-Regimente, Graf Ratibor von Werßowitz zu Potsdam

```

<PER.ind>
  <COMP.title>Königl. Preußische Lieutenant im
    <ORG.ent>Garde-Uhlanen-Regimente </ORG.ent>
  </COMP.title>
  ,
  <COMP.title>Graf</COMP.title>
  <COMP.name>Ratibor</COMP.name>
  <COMP.demonym>von Werßowitz</COMP.demonym>
  <COMP.demonym>zu Potsdam</COMP.demonym>
</PER.ind>

```

ehemalige Gouverneur von Catalonien, Graf Espagne

```

<PER.ind>
  <COMP.qualifier>ehemalige</COMP.qualifier>
  <COMP.func>Gouverneur</COMP.func>
  <COMP.demonym>von Catalonien</COMP.demonym>
  ,
  <COMP.title>Graf</COMP.title>
  <COMP.name>Espagne</COMP.name>
</PER.ind>

```

Sr. hochfürstlichen Durchl. des Prinzen Friedrich, Sohnes Sr. Hoheit des Kurprinzen.

```

<PER.ind>

```

```

    <COMP.title>Sr. hochfürftlichen Durchl.</COMP.title>
    des
    <COMP.title>Prinzen</COMP.title>
    <COMP.name>Friedrich</COMP.name>,
    <COMP.func>Sohnes
      <PER.ind>
        <COMP.title>Sr. Hoheit</COMP.title>
        des
        <COMP.title>Kurprinzen </COMP.title>
      </PER.ind>
    </COMP.func>
  </PER.ind>

```

Kaiferlich Oeſterreichſche Kabinets-Kourier Vardioro.

```

  <PER.ind>
    <COMP.title>Kaiferlich</COMP.title>
    <COMP.demonym>Oeſterreichſche</COMP.demonym>
    <COMP.func>Kabinets-Kourier</COMP.func>
    <COMP.name>Vardioro</COMP.name>
  </PER.ind>

```

Prinzeßinen Töchtern, Amalie und Maria.

```

  <PER.coll>
    <COMP.title>Prinzeßinen</COMP.title>
    Töchtern,
    <COMP.name>Amalie</COMP.name>
    und
    <COMP.name>Maria</COMP.name>
  </PER.coll>

```

Kommandeurs des Ordens des Heiligen Geiſtes den Kardinal de la Luzerne, den Kardinal de Bausset, den Erzbifchof von Bordeaux und den Abbe Montesquiou.

```

  <PER.coll>
    <COMP.func>Kommandeurs des
      <ORG.ent>Ordens des Heiligen Geiſtes</ORG.ent>
    </COMP.func>
    den
    <PER.ind>
      <COMP.func>Kardinal</COMP.func>
      de la
      <COMP.demonym>Luzerne</COMP.demonym>
    </PER.ind>,
    den
    <PER.ind>
      <COMP.func>Kardinal</COMP.func>
      <COMP.name>de Bausset</COMP.name>
    </PER.ind>,
  </PER.coll>

```



```

den
  <COMP.func>Erzbischof</COMP.func>
  <COMP.demonym>von Bordeaux</COMP.demonym>
</PER.ind>
und
<PER.ind>
  den
    <COMP.func>Abbe</COMP.func>
    Montesquiou
  </PER.ind>
</PER.coll>

```

Herzog und die Herzogin von Angouleme

```

<PER.coll>
  <COMP.title>Herzog</COMP.title>
  und die
    <COMP.title>Herzogin</COMP.title>
    <COMP.name>von Angouleme</COMP.name>
</PER.coll>

```

Frau von Lepel, geb. v. d. Lanken

```

<PER.ind> Frau von
  <COMP.name> Lepel </COMP.name>
  , geb. v. d.
  <COMP.name> Lanken </COMP.name>
</PER.ind>

```

Staatssekretair und Chefpräsident der Hauptbank Friefe

```

<PER.ind>
  <COMP.func>Staatssekretair</COMP.func>
  und
  <COMP.func> Chefpräsident der
    <ORG.adm>Hauptbank</ORG.adm>
  </COMP.func>
  <COMP.name> Friefe</COMP.name>
</PER.ind>

```

Nested entities with components are also annotated:

Se. Exc. der Generallieutenant, diesseitiger außerordentlicher Gefandter und bevollmächtigter Minister am Rußisch Kaiserlichen Hofe Freiherr von Schöler.

```

<PER.ind>
  <COMP.title>Se. Exc. der Generallieutenant</COMP.title>,
  <COMP.func>diesseitiger außerordentlicher Gefandter</COMP.func>
  und
  <COMP.func>bevollmächtigter Minister am
    <PER.coll>Rußisch Kaiserlichen Hofe</PER.coll>
  </COMP.func>

```

<COMP.title>Freiherr</COMP.title>  
 <COMP.name>von Schöler</COMP.name>  
 </PER.ind>

Britische General-Konful in Tripolis, Herr Warrington

<PER.ind>  
 <COMP.func>  
 <COMP.demonym>Britische</COMP.demonym>  
 General-Konful in  
 <LOC.adm>Tripolis</LOC.adm>  
 </COMP.func>  
 ,Herr  
 <COMP.name>Warrington</COMP.name>  
 </PER.ind>

Nangesdorf, Sec. Lt. von demf. Regt.

<PER.ind>  
 <COMP.name>Nangesdorf</COMP.name>,  
 <COMP.title>Sec. Lt.</COMP.title>  
 von demf. Regt.  
 </PER.ind>

Professor der Theologie an der Universität Würzburg, Dr. Stein

<PER.ind>  
 <COMP.title>Professor der Theologie an der Universität Würzburg </COMP.title>,  
 <COMP.title>Dr.</COMP.title>  
 <COMP.name>Stein </COMP.name>  
 </PER.ind>

Gemahlin des Geheimen Kommerzien-Raths Henschel zu Kassel

<PER.ind>  
 <COMP.func>Gemahlin</COMP.func>  
 des  
 <PER.ind>  
 <COMP.func>Geheimen Kommerzien-Raths</COMP.func>  
 <COMP.name>Henschel</COMP.name>  
 <COMP.demonym>zu Kassel</COMP.demonym>  
 </PER.ind>  
 </PER.ind>

verehelichte Klempnermeister Jasch, Louise, geborne Berger, zu Nieder-Hermsdorf

<PER.ind>  
 <COMP.func>verehelichte  
 <PER.ind>  
 <COMP.func>Klempnermeister</COMP.func>  
 <COMP.name>Jasch</COMP.name>  
 </PER.ind>  
 </COMP.func> ,

<COMP.name>Louise</COMP.name>,  
 geborne <COMP.name>Berger</COMP.name>,  
 <COMP.demonym>zu Nieder-Hermsdorf</COMP.demonym>  
 </PER.ind>

königlich schwedischer Major Nyguist im Generalstabe

<PER.ind>  
 <COMP.title>königlich schwedischer Major </COMP.title>  
 <COMP.name> Nyguist <COMP.name>  
 </PER.ind>

im Generalstabe

-> if Generalstabe were after title it would have been included in title

Gefchwifter Laura, Amalie und Gifela Neumann

<PER.coll>  
 <COMP.func>Gefchwifter</COMP.func>  
 <PER.ind>Laura</PER.ind>,  
 <PER.ind>Amalie</PER.ind>  
 und  
 <PER.ind> Gifela</PER.ind>  
 <COMP.name> Neumann</COMP.name>  
 </PER.coll>

Oberförfter Joly zu Rahm bei Großenbaum, Reg. Bez. Düffeldorf

<PER.ind>  
 <COMP.func>Oberförfter</COMP.func>  
 <COMP.name>Joly</COMP.name>  
 <COMP.demonym>zu Rahm</COMP.demonym>  
 </PER.ind>

bei

<LOC.adm>Großenbaum</LOC.adm>,  
 <LOC.adm>Reg. Bez. Düffeldorf</LOC.adm>

Kaufleute G. Rosenthal

<PER.coll>  
 <COMP.func> Kaufleute </COMP.func>  
 <ORG.ent><PER.ind> G. Rosenthal </PER.ind> & Co.</ORG.ent>  
 </PER.coll>

Gebr. Jückstock

<PER.coll>  
 Gebr.  
 <COMP.name>Jückstock</COMP.name>  
 </PER.coll>

Firma J. Tifcher Nachfolger (Gebr. Gerich)

```

<ORG.ent>Firma
  <PER.ind>J. Tifcher</PER.ind>
</ORG.ent> Nachfolger
(<PER.coll>Gebr. Gerich</PER.coll>)
Herren H.F. Fetschow & Sohn
<PER.coll>
  <COMP.title> Herren </COMP.title>
  <ORG.ent>
    <PER.ind>H.F. Fetschow </PER.ind>
    & Sohn</ORG.ent>
</PER.coll>
Russischer Hof
<PER.coll>Russisch Kaiserlicher Hof </PER.coll>
How to choose between COMP.demonym and LOC.adm?

```

**questions for person (will be revised by lillian)** Page 13 in [1].

How to treat the following cases? \* surnames with “von” and a city \* surnames with “Abbe” and a city \* etc

How do we choose between these cases? \* a person contains COMP.func with LOC.adm inside

```

<PER.ind>
  <COMP.func>Abbe
    <LOC.adm>Montesquiou</LOC.adm>
  </COMP.func>
</PER.ind>

```

- a person contains COMP.func with COMP.demonym inside:

```

<PER.ind>
  <COMP.func>Abbe
    <COMP.demonym>Montesquiou</COMP.demonym>
  </COMP.func>
</PER.ind>

```

- a person contains COMP.func only

```

<PER.ind>
  <COMP.func>Abbe Montesquiou</COMP.func>
</PER.ind>

```

- a person contains COMP.func and COMP.demonym

```

<PER.ind>
  <COMP.func>Abbe</COMP.func>
  <COMP.demonym>Montesquiou</COMP.demonym>
</PER.ind>

```

- a person contains COMP.func and LOC.adm

```
<PER.ind>
  <COMP.func>Abbe</COMP.func>
  <LOC.adm>Montesquiou</LOC.adm>
</PER.ind>
```

Similar cases:

```
<PER.ind>
  <COMP.func>Kardinal</COMP.func>
  de
  <COMP.demonym> la Luzerne</COMP.demonym>
</PER.ind>

<PER.ind>
  den
  <COMP.func>Erzbischof </COMP.func>
  <COMP.demonym>von Bordeaux</COMP.demonym>
</PER.ind>
```

## Organization (ORG)

### Subtypes

- ORG.adm: an organisation which plays a mainly administrative role.
- ORG.ent: an organisation which doesn't play a mainly administrative role

### Coverage of the type Organization

- Considered as Organization:
  - registered organizations
  - museums, institutes, universities, libraries (annotate as ORG.ent)
  - restaurants (annotate as ORG.ent)
  - military units (annotate as ORG.ent)
  - political parties (annotate as ORG.ent)
  - firms (annotate as ORG.ent)
  - hospitals (annotate as ORG.ent)
  - factories (annotated as ORG.ent, e.g., vereinigten rheinisch-westfälische Pulverfabriken, Confolidirtes Braunkohlen-Bergwerk Marie bei Atzendorf)
  - railways (annotated as ORG.ent, e.g., Kerkerbachbahn)
  - press agencies (annotate as ORG.ent, e.g., W.T.B.)
  - exchanges (annotate as ORG.ent but not linked, e.g., Berliner Börse)
  - chamber of commerce (annotate as ORG.ent)
  - armies of a country (annotate as ORG.ent, e.g., Preußische Armee)
  - if something is both PER.coll and ORG.ent both entities are added
  - parliaments (annotate as ORG.adm, e.g., Reichstag, Unterhaus)
  - governments if it is known what they govern (annotated as ORG.adm, e.g., französische Regierung, Regierung von Köln)

- ministries (annotated as ORG.adm)
- military authorities (Militärbehörde) (annotate as ORG.adm)
- courts of justice (annotate as ORG.adm)
- public prosecution service (Staatsanwaltschaft) (annotate as ORG.adm)
- central banks (Zentralbank, Reichsbank) (annotate as ORG.adm)
- departments if the organization to which they belong is known (e.g., Militär-Medizinal-Abtheilung des Kriegsministeriums)
- former names of an organization are added (vorm.)
- organisations that are distinctive (exist only once) are annotated without the location
- a location is added to the annotation if it is part of the name (for ex. Königlichen Eifenbahn-Direktion Berlin)
- Not considered as Organization:
  - theaters (annotate it as LOC.fac)
  - departments if the organization to which they belong is unknown (e.g., Abtheilung des Innern)
  - governments if it is unknown what they govern (annotated as ORG.adm, e.g., Kaiserliche Regierung, Regierung in/zu Cöln)

### specific heuristics

- Bundesrat, Bern are seperately annotated
- Königliche Regierungs-Hauptkasse, Königlichen Staatsschulden-Tilgungskasse etc. (annotate as ORG.adm)
- banks that are named after a person (anntoate as ORG.ent & PER.ind or PER.coll)
- Reserve (milit.) (annotate as ORG.ent)
- other organisations: Police, roman church, schools, international committees, orphanages are annotated as well
- mail system:
  - Postanstalten, Postämter are annotated as ORG.ent if they are specific (e.g., General-Postamt/ Reichspostamt)
  - Postverwaltung is annotated as ORG.adm
  - Post is annotated as ORG.ent
- Not considered as Organization:
  - Köngl. Preußische Lotterie not annotated
  - Regierung in/zu Cöln not annotated

### Tricky cases for Organization Blücherfchen Heere

```

<ORG.ent>
  <PER.ind>Blücherfchen</PER.ind>
  Heere
</ORG.ent>

<PER.coll>
  <COMP.title> Herren </COMP.title>
  <ORG.ent>
    <PER.ind>H.F. Fetschow </PER.ind>& Sohn

```

</ORG.ent>  
 </PER.coll>  
 Pariser Bijoutier Odiot  
 <ORG.ent>  
     Pariser  
     Bijoutier  
     <PER.ind>Odiot</PER.ind>  
 </ORG.ent>  
 PER.ind not annotated  
 Niederländische Regierung  
 <ORG.adm>niederländische Regierung</ORG.adm>  
 Justizkommission des Reichstages  
 <ORG.adm>Justizkommission des Reichstages</ORG.adm>  
 Militär-Medizinal-Abtheilung des Kriegsministeriums  
 <ORG.adm> Militär-Medizinal-Abtheilung des  
     <ORG.adm> Kriegsministeriums </ORG.adm>  
 </ORG.adm>  
 Königliches Handelsgericht-Sekretariat  
 <ORG.adm>Königliches Handelsgericht-Sekretariat</ORG.adm>  
 Firma: Philipp Cullmann  
 <ORG.ent> Firma:  
     <PER.ind> Philipp Cullmann </PER.ind>  
 </ORG.ent>  
 Lehranstalt der Jefuiten in Sikawei  
 <ORG.ent>Lehranstalt der Jefuiten in Sikawei</ORG.ent>  
 Hôtel Zulaft in Mülheim a. d. Ruhr  
 <ORG.ent>Hôtel Zulaft</ORG.ent> in Mülheim a. d. Ruhr  
 Firma J. Tifcher Nachfolger Gebr. Gerich  
     <ORG.ent>Firma  
         <PER.ind>J. Tifcher</PER.ind>  
 </ORG.ent> Nachfolger  
 (<PER.coll>Gebr. Gerich</PER.coll>)  
 Actien-Gesellschaft für Eisen-Industrie und Brückenbau vorm. J. C. Harkort  
 <ORG.ent>Actien-Gesellschaft für Eisen-Industrie und Brückenbau vorm.  
     <PER.ind>J. C. Harkort</PER.ind>  
 </ORG.ent>

## **Location (LOC)**

### **Subtypes**

- LOC.adm: a territory with a geopolitical border (e.g., cities, countries & continents)
- LOC.fac: named buildings (train station & museum) & their extensions (stadium & campus); a physical location of an organisation
- LOC.oro: an oronym (e.g., are streets, squares, roads & highways)
- LOC.phys: a physical location (e.g., mountains, rivers & planets)

### **Coverage of the type Location**

- Considered as Location:
- Considered as LOC.adm:
  - ancient cities (e.g. Pompeii)
  - Islands
  - administrative areas of the church (e.g., Diözese)
  - if a state is named with the city (e.g., Königsberg (Ostpr.)) both are annotated individually
  - the discription of the administrative region (e.g., Gemeinde, Kreise) is included in the annotation
  - if a river is named with the city (e.g., Ludwigshafen, Rhein) the river is included in annotation
- considered LOC.fac:
  - specific buildings: military buildings, theaters etc. (e.g., Schloss Mannheim, Theaters, train stations)
  - building that only exist once in a city (e.g., Stadt-Theater in Berlin) are annotated with the following city
  - unspecific buildings (e.g., Haus)
- considered LOC.phys:
  - Gardens and (national)parcs (e.g. Jardin des Tuileries)
  - specific river banks (e.g. rechtes Rhein Ufer)
  - valleys (e.g., Moselthal)
- Not considered as LOC.adm:
  - regions without clear borders (e.g., Ostafrika, Ostindien)
  - if a location is written as an adjective (e.g., Spanisch)
  - colonies that can not be attributed to a specific geographical location (e.g., Spanifchen Kolonien)
- Not considered as LOC.add:
  - street names without a house number
- Not considered as LOC.fac:
  - ports
- Specifications (e.g., island, colony etc.) of locations are annotated as well
- In case of locations that can't be clealy attributed to a city or another location because they are both a city and a location with the same name exist (for example: Samos -> is both city and island) no QID is assigned
- if two locations are separable, they are annotated individually



- if two locations are not separable, they are annotated together (e.g., Rahm-Reg.-Bez. Düsseldorf)

### specific heuristics

- Ireland, when alone: <https://www.wikidata.org/wiki/Q22890>
- Great Britain and Ireland, together (1801- 1927): <https://www.wikidata.org/wiki/Q174193>
- Great Britain, when alone: <https://www.wikidata.org/wiki/Q23666>
- Berl n / Leipz g to Berlin / Leipzig
- Rio is not annotated as Rio de Janerio
- Gefängnis is not annotated
- LOC.fac: Embassies are annotated if it is clear where they are and in which country they originate
- LOC.fac: Stadt-Theater in city x are annotated with the following city, Stadt-Theater without a city are annotated alone, same procedure with Church in city x and branches of banks (e.g., Reichsbankstelle in...)
- Yellowstone / Big Horn: LOC.phys without identifier
- West indies is a Danish Colony in the Caribbean: annotated
- Westaustralien is a state in Australia: annotated
- Windsor is annotated as LOC.adm
- El Salvador is annotated as LOC.adm
- Vereinigte Staaten von Nordamerika, Etats Unis (franz. Name) is annotated as USA
- Vereinigte Staaten are not annotated
- Goldküste is annotated as LOC.adm
- Rio is not annotated as Rio de Janerio

### Tricky cases for Location Kolonie Bourbon

<LOC.adm>Kolonie Bourbon</LOC.adm>

Königsberg

<LOC.adm>Königsberg</LOC.adm>

Königsberg in Pr.

<LOC.adm>Königsberg in Pr.</LOC.adm>

Berlin Breitestaße No 20

<loc.adm> Berlin </loc.adm>

<loc.add> Breitestaße No 20 </loc.add>

Garnison Mühlberg

<LOC.fac> Garnison Mühlberg</LOC.fac>

### EVENT

- Schlacht an der Katzbach: specific (specific in space and time)
- 2. Kongress der russischen Ärzte: specific

- Geburt des jungen Prinzen: not specific

## **DATE**

- Ein u. Zwanzigsten März 1836 -> annotated
- 21ster d. Mts -> not specific -> not annotated
- Ende Dezember 1831: not specific -> not annotated
- Dezember 1831: specific -> annotated
- 1876/77: only 1876 -> annotated/ identifier added
- 1. Juli 1837 bis Ende Dezember 1838: annotated as TIME.range and TIME.date.abs is added to individual dates

## **specific heuristics**

- Michaelis d. J. -> 29. September

## **Tricky cases for DATE**

1. Juli 1837 bis Ende Dezember 1838

```
<TIME.range>
  <TIME.date.abs>1. Juli 1837</TIME.date.abs>
  bis
  <TIME.date.abs>Ende Dezember 1838</TIME.date.abs>
</TIME.range>
```

## **PROD.media**

names Preußischer Staatsanzeiger: <https://www.wikidata.org/wiki/Q55257346>

There are multiple names for the Reichsanzeiger newspaper (https://digi.bib.uni-mannheim.de/periodika/reichsanzeiger/).

## **specific heuristics**

- Bundes-Gesetzblatt is annotated as PROD.media

## **Entity linking**

- entities are linked against Wikidata
- nested entities are linked unless the main entity is PER.ind
- All types of entities except for components and TIME.range are linked
- LOC.adm: only linked to address if a Wikipedia article with the exact name of the address exists (in our dataset no such Wikipedia articles existed)
- If the historical referent differs from the current referent (e.g., Reichsgaue Sudetenland) the historical Wikidata entry is linked. If there is no historical referent, the current Wikidata entry is linked.

- different iterations of the same organisation (e.g., the Reichstag) are not linked to the specific iteration (e.g., 10. Reichstag) but the general wikidata entry (Reichstag - Wikidata)
- abbreviations are linked as well if they can be clearly attributed (e.g., SS)
- in case of a metonymy of PER.coll, PER.ind and ORG.ent (e.g., Firma Gebrüder F. J. Badart -> both ORG.ent and PER.coll) the entities are not linked because distinction is too difficult
- parties with generic names (e.g., liberale Partei) are not linked
- if a LOC.adm is both city and district we use the identifier of the city unless it is known that the district is meant
- for lists of administrative areas (e.g., Regierungsbezirken Arnsberg, Düsseldorf, Potsdam, Hamburg) the identifier of the administrative area in the beginning (e.g., Regierungsbezirk) is used

### **specific heuristics**

- argentinische Republik is linked
- Senegambien linked as British Senegambien
- Puerto Rico is not linked
- Tokio before 1889 is not linked
- Tokio after 1889 is linked as Tokio - Wikidata
- Römisches Reich is linked as Ancient Rome - Wikidata
- Malta is linked as Island of Malta - Wikidata
- Württemberg from 1806 to 1918: Kingdom of Württemberg - Wikidata
- Nordamerika -> Kontinent is linked as Nordamerika - Wikidata
- Goldküste is annotated because its a LOC.adm but no identifier because its not clear which LOC.adm is meant
- Cordoba is not linked

### **References**

- [1] Ehrmann, Watter, Romanello, & Clematide. (2019). Impresso Named Entity Annotation Guidelines (2.1). Zenodo. <https://doi.org/10.5281/zenodo.3585750>
- [2] Romanello, Matteo, & Najem-Meyer, Sven. (2022). Guidelines for the Annotation of Named Entities in the Domain of Classics (18.03.2022). Zenodo. <https://doi.org/10.5281/zenodo.6368101>
- [3] Ahmed Hamdi, Elvys Linhares Pontes, & Antoine Doucet. (2021). Annotation Guidelines for Named Entity Recognition, Entity Linking and Stance Detection (v3.1). Zenodo. <https://doi.org/10.5281/zenodo.4574199>
- [4] Menzel, Sina, Zinck, Josefine, & Petras, Vivien. (2020). Guidelines for Full Text Annotations in the SoNAR (IDH) Corpus (2.4). Zenodo. <https://doi.org/10.5281/zenodo.5116015>
- [5] Sophie Rosset, Cyril Grouin & Pierre Zweigenbaum. (2011). Entités nommées structurées: guide d'annotation Quaero. [pdf]
- [6] Cyril Grouin, Sophie Rosset, Pierre Zweigenbaum, Karén Fort, Olivier Galibert, and

Ludovic Quintard. (2011). Proposal for an Extension of Traditional Named Entities: From Guidelines to Evaluation, an Overview. In Proceedings of the 5th Linguistic Annotation Workshop, pages 92–100, Portland, Oregon, USA. Association for Computational Linguistics. [pdf]