

# GerPS-NER: A Dataset for Named Entity Recognition to Support Public Service Process Creation in Germany

Leila Feddoul<sup>1</sup>, Sarah T. Bachinger<sup>1,2,\*</sup>, Clara Lachenmaier<sup>3</sup>, Sebastian Apel<sup>4</sup>, Pirmin Karg<sup>1</sup>, Norman Klewer<sup>1</sup>, Denys Forshayt<sup>1</sup>, Robin Erd<sup>1</sup> and Marianne Mauch<sup>1,2</sup>

<sup>1</sup>Heinz Nixdorf Chair for Distributed Information Systems, Friedrich Schiller University Jena, Germany

<sup>2</sup>Competence Center Digital Research (zedif), Friedrich Schiller University Jena, Germany

<sup>3</sup>Computational Linguistics, Bielefeld University, Germany

<sup>4</sup>City Administration Jena, Germany

## Abstract

For the end-to-end digitization of public administration in Germany, reliable entity recognition in legal texts is essential to support the creation of corresponding processes (legal norm analysis). To the best of our knowledge, datasets that can serve to train and test machine learning models on this specific entity recognition task with our domain-specific categories do not exist. Therefore, we present *GerPS-NER*, a dataset for entity recognition to support the legal norm analysis, consisting of 24k sentences from German law documents annotated with ten categories (e.g., main actor). We showcase the dataset generation workflow, including the data collection, the annotation guidelines, and the annotation phases. The dataset [1] is publicly available under a CC-BY 4.0 license.

## Keywords

Dataset, Named Entity Recognition, Legal Norm Analysis, Annotation, Machine Learning, Public Administration

## 1. Introduction

Public administration institutions follow processes based on legislation (e.g., laws and ordinances), when delivering a public service like car registration to citizens or companies. In Germany, the Federal Information Management (FIM) [2] provides standardized methods for analyzing such legal bases. Trained public administration employees mark relevant terms or sentences of categories, such as the main actor as well as the result receiver, and create a list of

---

*Second International Workshop on Semantic Technologies and Deep Learning Models for Scientific, Technical and Legal Data SemTech4STLD, ESWC '24, May 26th or 27th, 2024, Hersonissos, Greece*

\*Corresponding author.

✉ leila.feddoul@uni-jena.de (L. Feddoul); sarah.bachinger@uni-jena.de (S. T. Bachinger); marianne.mauch@uni-jena.de (M. Mauch)

🌐 <https://www.fusion.uni-jena.de/people/details/leila-feddoul> (L. Feddoul);

<https://www.zedif.uni-jena.de/de/team/sarah-bachinger.html> (S. T. Bachinger)

🆔 0000-0001-8896-8208 (L. Feddoul); 0009-0005-5422-2164 (S. T. Bachinger); 0000-0002-9207-3420

(C. Lachenmaier); 0009-0008-5711-5390 (P. Karg); 0009-0000-9755-8892 (N. Klewer); 0000-0001-5340-8163

(R. Erd); 0000-0003-1478-1867 (M. Mauch)



© 2024 Copyright for this paper by its authors. Use permitted under Creative Commons License Attribution 4.0 International (CC BY 4.0).

discovered process steps together with the related data fields<sup>1</sup>. Finally, they assemble all these elements in a process using a restricted and an extended BPMN [3] notation. This process is used to create web forms allowing an online delivery of the service for potential applicants. In this way, the service, the associated process, and data fields matching specific process steps are described on the basis of the associated legal basis. This annotation of legal texts with categories is called the FIM – legal norm analysis, which represents the first step in the creation of administrative services with the identification of all involved process elements (e.g., steps or actors).

The Canarèno [4] (“Computer-assisted analysis of electronically available legal norms”) project is one of three research projects of the “Open Design of Digital Administrative Architectures (openDVA)”<sup>2</sup> working group investigating the path from the legal text to its digital implementation and how it can be simplified and partially automated, both for new or existing legal texts. The project aims to support this manual and time-consuming FIM analysis of German legal norms by automatically generating suggestions that assign categories<sup>3</sup> to relevant terms/sentences, allowing users to review them (accept, edit, or delete), and capturing corrections to continuously improve the system (Human-in-the-Loop [5]). We aim to leverage techniques from natural language processing, such as Named Entity Recognition (NER) [6], for the category detection. For this purpose, we investigate different methods for solving it. Supervised machine learning approaches are promising, but some methods need a relatively large amount of annotated training data. Since such data is not available, we create the GerPS-NER corpus, which can serve as a training and evaluation sets for any model aiming to annotate German legal texts with our categories. In the following, we focus on explaining the workflow for the creation of this corpus. Our contributions can be summarized as follows:

- Describing the dataset creation workflow including our iterative approach for designing annotation guidelines.
- Proposing GerPS-NER, a dataset consisting of 24k annotated sentences with ten categories of the legal norm analysis.

The dataset [1] is publicly available under a CC-BY 4.0 licence, together with the code used for law crawling, data processing, and the conversion of the annotated corpus into the final standard format (IOB2 [7]).

## 2. Related Work

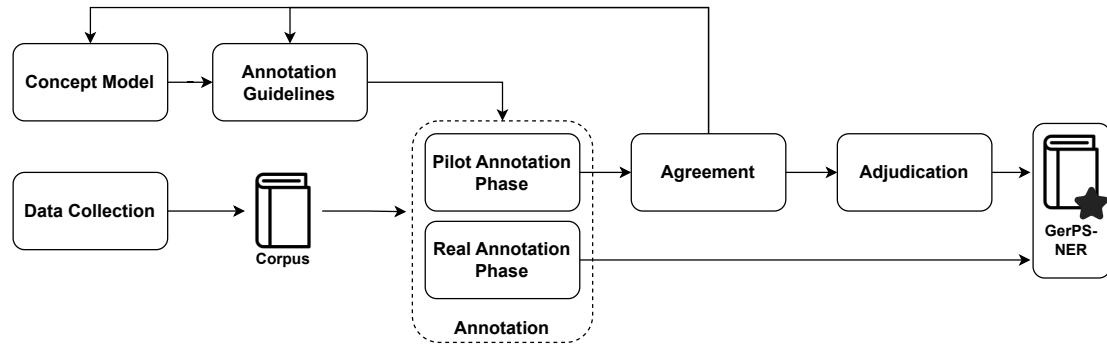
Some datasets for NER on German legal texts exist, e.g., Leitner et al. [8] created a dataset using 19 entities (e.g., *judge*) for annotating German court decisions, called LER [9]. Darji et al. [10] published a dataset of legal references in German law that were annotated (e.g. *Buch* (book)) by law experts. Wrzalik and Krechel [11] created a dataset of case documents and describe labelled queries to the open legal platform Open Legal Data for German Legal Information Retrieval

---

<sup>1</sup>Data fields represent the different pieces of information of online forms (e.g., first name, profession) that are necessary for the result receiver to apply for a service.

<sup>2</sup>[www.opendva.de](http://www.opendva.de)

<sup>3</sup>Refer to Table 1 for the definitions



**Figure 1:** Workflow for creating GerPS-NER

called GerDaLIR. To the best of our knowledge, there does not exist a corpus that is annotated with the categories needed for legal norm analysis on German legal texts.

### 3. Dataset Creation Workflow

The pipeline for creating the training corpus is depicted in Figure 1 and is based on [12]. First, a conceptual model was defined to set the categories used for annotation and their potential relations (refer to [13, 14] for more details about the created ontology). This was followed by collecting relevant text data to construct the corpus and a first draft of the annotation guidelines (refer to Section 4). After instructing annotators, the actual annotation phase started. We divided this into: (1) A pilot annotation phase with a small number of documents where the agreement between the different annotators (Inter-Annotator Agreement (IAA)<sup>4</sup>) is considered after each iteration. Here, the task definitions and guidelines are evaluated and improved based on annotators’ feedback. In an adjudication step, annotation mismatches are resolved and a first annotated corpus (gold standard) is created. (2) A real annotation phase where the improved guidelines are applied but not fundamentally changed anymore. Each remaining document is annotated only by a single annotator, omitting (IAA).

#### 3.1. Concept Model and Data Collection

The categories used for GerPS-NER were based on those defined by FIM - legal norm analysis and were extended during the project after agreement with the experts to allow a more concise process description. A basic definition of the categories is given in Table 1, for a more detailed description, refer to the latest annotation guideline<sup>5</sup>.

After the definition and semantic description of the categories that should be detected in the law texts, relevant data needs to be collected. For this purpose, we gather law paragraphs that are used as a basis for the creation of public service processes in Germany. The website

<sup>4</sup>A measure for the overlap between two annotators, e.g. Cohen’s kappa [15].

<sup>5</sup>[1] in folder `annotation_guidelines/10-01-2024_Annotation_Guide_V4.pdf`

**Table 1**

Category definitions with examples in German

Category	Definition	Example
<i>Hauptakteur</i> (main actor)	From public administration view, the person or office that is responsible for the service.	Bundesamt für Sicherheit in der Informationstechnik
<i>Ergebnisempfänger</i> (result receiver)	Person or company that wants to use the service	Antragsteller
<i>Mitwirkender</i> (contributor)	External office or another role that is taken into account	Bundesministerium des Inneren
<i>Aktion</i> (action)	Action of public administration or other actors	beantragen, vorlegen, erteilen
<i>Signalwort</i> (signal word)	Level of commitment; expressed by modal verbs or other descriptive terms	bei Bedarf, erforderlich, zweckgebunden, auf Verlangen, muss, soll, kann
<i>Dokument</i> (document)	Documents that are exchanged between actors	Sicherheitszertifikat, Antrag
<i>Bedingung</i> (condition)	Requirement for an action (e.g. a temporal component); certain words that start a conditional clause	Die Anträge werden in der zeitlichen Reihenfolge ihres Eingangs bearbeitet; hiervon kann abgewichen werden, [wenn das Bundesamt wegen der Zahl und des Umfangs anhängiger Prüfungsverfahren eine Prüfung in angemessener Zeit nicht durchführen kann und an der Erteilung eines Zertifikats ein öffentliches Interesse besteht].
<i>Frist</i> (deadline)	Deadline or a timelimit that are relevant for the process	innerhalb von 6 Wochen
<i>Datenfeld</i> (data field)	Explicit naming of entities that are relevant for a form	In der Anzeige sind anzugeben: 1. der [Zweck des Eingriffs], .... und voraussichtliche [Dauer des Vorhabens].
<i>Handlungsgrundlage</i> (legal basis)	Cross reference to other laws	Absatz 3, Satz 2 Absatz 2 Nr. 3

*FIM-Portal*<sup>6</sup> gives an overview of available services where, if available, links to the relevant law paragraphs are accessible via the “Rechtsgrundlage” field of each service.

We started with an internal initial pilot phase involving three human annotators and consisting of 10 documents that were randomly selected from data we collected from the *FIM-Portal*. The analysis of distribution of the collected service law paragraphs using the first version of the data collection code revealed a non-uniform distribution of the services with respect to their groups<sup>7</sup> (e.g., health) meaning that some services were more often represented than others. In addition, we noticed that we cover only 33 of the 160 available service groups (21%). For this reason in the real annotation phase, we collected data using another data collection strategy which covers all the groups in a relatively uniform way and thus creates a diverse and balanced training corpus. To create the latter, a list of services provided by the *FIM-Portal* is used as input. This list is given as a CSV file<sup>8</sup> including the content of the corresponding HTML pages. This file can be downloaded after registration from the internal section of the *FIM-Portal* catalogue [16]

<sup>6</sup><https://fimportal.de/>

<sup>7</sup>Refer to [https://www.xrepository.de/api/xrepository/urn:de:fim:leika:leistungsgruppierung\\_20231229/download/FIMLeiKaLeistungsgruppierungen\\_20231229.xlsx](https://www.xrepository.de/api/xrepository/urn:de:fim:leika:leistungsgruppierung_20231229/download/FIMLeiKaLeistungsgruppierungen_20231229.xlsx) for the different service groups.

<sup>8</sup>Downloads\_im\_CSV-Format\_LeiKa-plus\_Alle\_Leistungen\_inkl\_inhaltlicher\_Beschreibungen\_mit\_HTML.csv

and contains the links to law paragraphs corresponding to the different services. Only links pointing to a paragraph on the website *gesetze-im-internet.de*<sup>9</sup> page are considered<sup>10</sup>. For each of the those links, we crawl the content from *gesetze-im-Internet*. We select a fixed number (10) of services from each group type. Each collection of paragraphs related to a specific service are stored in a separate document and identified using the service ID (LeiKa [2]). Ideally, one would have 10 documents of each type of service, but for some there were not enough, so for these types of service there are less documents. In total, we collected 1020 documents from 141 service types. Note that we first removed the 10 already annotated documents during the previous initial pilot phase which means that we had 1010 documents available for the real annotation phase. The code for data crawling [17] and the balanced corpus without annotations [18] are published on Zenodo under an CC BY 4.0 licence.

## 3.2. Annotation Phase

With the aim of iteratively creating an annotation guideline for a larger annotation campaign, we started an initial pilot phase with a set of 10 documents. The annotation was performed by three annotators using the annotation tool INCEpTION [19], followed by a calculation of the IAA, an adjudication step to solve mismatches, and a discussion of open issues with domain experts. In the following, we will give more details about the adjudication phase of the initial pilot phase and the real annotation phase.

### 3.2.1. Adjudication (Initial pilot phase)

The process for adjudication and thus the consolidation of the annotations done by all three annotators to one annotated gold standard document was performed as follows using the curator (adjudicator) interface of INCEpTION [19]:

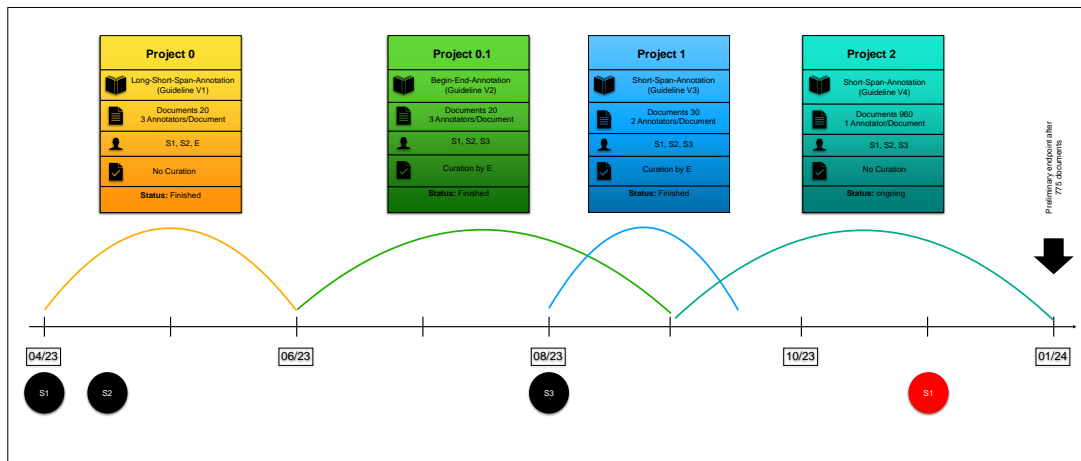
- Consider only the lines with annotations.
- Automatically accept annotations where more than one annotator agree.
- If only two out of three annotators annotated a sentence (meaning the third judged the sentence not relevant), we consider the annotations that were marked.
- Correct annotation based on the newest version of the annotation guidelines.
- For the annotations where less than 2 people agree (each one of the three annotators has annotated differently):
  - Try to converge and agree on one annotation.
  - If not possible, gather as an open issue to discuss with the expert. We grouped open questions per document<sup>11</sup>.
- If, during the process, possible guideline extensions arise, we write them down and add to the guidelines in the corresponding place.

---

<sup>9</sup>[www.gesetze-im-internet.de](http://www.gesetze-im-internet.de)

<sup>10</sup>E.g., [https://www.gesetze-im-internet.de/betrsvchv\\_2015/\\_19.html](https://www.gesetze-im-internet.de/betrsvchv_2015/_19.html) which links to paragraph 19 but not [https://www.gesetze-im-internet.de/betrsvchv\\_2015/](https://www.gesetze-im-internet.de/betrsvchv_2015/), which links to the entire law.

<sup>11</sup>[1] in folder adjudication/expert\_template.md



**Figure 2:** Phases of the real annotation phase: Here, the different projects of the annotation campaign are described. The annotation style (e.g. Long-Short-Span-Annotation) and annotation guideline are listed below the project name. Then, information about the number of annotated documents and the number of annotators per document is given. Finally, curation information and project status are given. The lines below the project descriptions show when they happened on a timeline from April 2023 to January 2024. The black circles indicate when the annotators started working on the annotation and the red ones when they left.

The guideline additions that originated from the adjudication phase are part of the “Annotation\_Guide\_V0.2”<sup>12</sup>.

The adjudication of the 10 documents of the initial pilot phase resulted in a first partial gold standard corpus. We provide a short<sup>13</sup> and an extended<sup>14</sup> version of the same corpus depending on the considered annotation span.

### 3.2.2. Real Annotation Phase

A larger annotation campaign involved four annotators (three law students (S1–3) and an employee in the municipal administration of city Jena (E)), as displayed in Figure 2. It also started with a pilot phase involving a small fraction of remaining documents. This is a training phase to ensure all new annotators have the same understanding of the developed annotation guideline and can annotate in a consistent way. The annotation campaign consists of 4 phases, each of which is organized as an individual project in INCEPTION [19]. Each annotator received an individual training at the beginning of their working time, which involved several individual meetings. The annotators were required to record questions, comments and the time they spent annotating for each annotated document. The individual phases are discussed in detail below.

<sup>12</sup>[1] in folder adjudication/curation\_guidelines\_addition.md

<sup>13</sup>[1] in folder intermediate\_corpora/adjudication/short

<sup>14</sup>[1] in folder intermediate\_corpora/adjudication/extended

**Project 0** With three annotators (S1, S2, and E), the first 20 documents were annotated according to the first version of the annotation guidelines<sup>15</sup>. The aim was to create two versions of the same corpus depending on the considered annotation span (long and short<sup>16</sup>). The weekly discussions revealed several points that called for action in the form of a revision of the annotation guidelines by mainly performing additions to the “general rules” and “specific rules” sections, that can be summarized as follows:

- Allow the annotation of nested occurrences (entity mentions embedded in longer entity mentions, e.g., a *condition* mentioning a required *document*).
- Annotate only the start and end of longer spans consisting of more than one word.
- If there is an “or” between words of the same type, (e.g., “Prüfung oder Befähigungsnachweise” (examination or certificates of qualification)), these should be annotated as single units.
- If there is a definition relevant to the service in the text, the entire definition is annotated using the category that is also used for the defined concept.
- If the title of the service description mentions “Intended for cancellation”<sup>17</sup> or something similar indicating this service is not up-to-date, the text should not be annotated.
- Negatively formulated actions or conditions, such as “Eine Approbation wird nicht erteilt” (A licence is not granted), must be marked as negative. To do this, there is a “Negation” checkbox, which already has “No” selected by default. If there is a negation, “Yes” must be entered by the annotators.

Each of the annotators annotated all the 20 documents. The latter were adjudicated (curated) by E, first just as a test round following the previously mentioned adjudication process of the initial pilot phase. This can be seen as a training for E, that will take over the adjudication in the next phase. Note that this phase corresponds to the first iteration (iteration 0) of the internal initial pilot phase as described in Section 3.2.1, which was basically a test phase, and the same documents will be annotated in the next iteration.

**Project 0.1** In this phase, the same first 20 documents were again annotated according to the second version of the annotation guidelines<sup>18</sup> with the same aim of the creation of two corpus versions. In this case, however, this was realized using a start-end annotation, where only the start and the end of relevant spans that consist of more than one word is annotated. The weekly discussions revealed that the annotation of long spans in particular led to problems because, despite the continuous expansion of the guidelines, the scope of what counts as long-span annotation could not be clearly defined. So the guidelines were again adapted to consider only short spans and the information about the annotation scope was added instead of all the sections that describe long span annotations:

As a rule, only one word is annotated per category. For example, let’s consider the *document* “Bescheinigung über die Wohnberechtigung” (certificate of entitlement to residence). Although

<sup>15</sup>[1] in folder [annotation\\_guidelines/30-03-2023\\_Annotation\\_Guide\\_V1.pdf](#)

<sup>16</sup>For the long span annotations, we consider more information than for the short span, refer to the afore mentioned annotation guideline for examples

<sup>17</sup>E.g., in German: “Vorgesehen zur Löschung”

<sup>18</sup>[1] in folder [annotation\\_guidelines/13-10-2023\\_Annotation\\_Guide\\_V2.pdf](#)



there is a phrase that describes the certificate in more detail, the core of the entire phrase and the word that can be considered as a document on its own is the certificate (“Bescheinigung”) and should therefore be annotated alone. Exceptions are the categories *Bedingung* (condition) and *Handlungsgrundlage* (legal basis). Each of the annotators annotated all the 20 documents. The documents were curated by E. The latter continued to follow the previous adjudication process and thus only intervened if there were fewer than two matches in an annotation. This project provides an intermediate gold standard with 20 documents<sup>19</sup>.

**Project 1** With the three annotators, 30 further documents were annotated according to the third version of the annotation guidelines<sup>20</sup>. From now on, only a short span version of the corpus with single-word annotations (except in exceptional cases) will be created. Each document was annotated by two of the annotators. For this purpose, each document was assigned one of the three combinations S1+S2, S1+S3 and S2+S3 to ensure that each person annotated equally often with the other two. The adjudication phase was performed by E. Here, the guideline underwent only minor changes such as: typos, removal of not needed rules<sup>21</sup>, extension/adaptation of some passages<sup>22</sup>, and the removal of the negation checkbox, because it was only rarely used. At the end, one document was removed<sup>23</sup>, because the legal basis corresponding to the service had changed. Thus the document was not annotated and the final number of annotated documents is 29. This project provides an intermediate gold standard with 29 documents<sup>24</sup>.

**Project 2** With the three annotators, the remaining 960 documents are being annotated according to the fourth version of the annotation guidelines<sup>25</sup>. Each document is now only processed by one annotator. Adjudication is therefore no longer necessary. From December 1st, 2023, annotation continued with only two annotators (S2 and S3). This project provides an intermediate gold standard with 775 of the 960 documents, that were annotated until January 31st, 2024<sup>26</sup>.

## 4. Annotation Guidelines

A crucial aspect in the creation of training corpora with human annotations are precise and comprehensive annotation guidelines. They define the task more precisely and aim to ensure consistent annotations by different annotators which is important when training models. For creating our annotation guidelines, the process is as follows:

### 1. Creation of a rudimentary set of guidelines

<sup>19</sup>[1] in folder code/intermediate\_corpora/annotation/Normenanalyse\_0.1

<sup>20</sup>[1] in folder annotation\_guidelines/13-10-2023\_Annotation\_Guide\_V3.pdf

<sup>21</sup>E.g., “If there is a definition relevant to the service in the text, the entire definition is annotated using the category that is also used for the defined concept”

<sup>22</sup>E.g., extended the explanation of the scope of annotation part

<sup>23</sup>a99040004076000.txt

<sup>24</sup>[1] in folder code/intermediate\_corpora/annotation/Normenanalyse\_1

<sup>25</sup>[1] in folder annotation\_guidelines/10-01-2024\_Annotation\_Guide\_V4.pdf

<sup>26</sup>[1] in folder code/intermediate\_corpora/annotation/Normenanalyse\_2



2. Annotation of a small number of documents with more than one annotator
3. Calculation of the IAA to show areas of high and low agreement
4. Discussion of issues and clarification
5. Refinement of the guidelines
6. Start again with Step 2

The initial pilot phase (10 documents from the unbalanced data collection) started by using a very basic and a short version of the guidelines “Annotation\_Guide\_V0”<sup>27</sup>. This first iteration (iteration 0) allowed the creation of a more detailed annotation guideline “Annotation\_Guide\_V0.1”<sup>28</sup>, that was used to annotated the same 10 documents again (iteration 1). This second iteration allowed to refine the annotation guideline again and to generate the first draft of the main version “Annotation\_Guide\_V0.2”<sup>29</sup> that consists of additions after the interview with the expert (a FIM coach) and the corresponding adjudication phase. This results in the first main version “Annotation\_Guide\_V1”<sup>30</sup> that is used in the real annotation phase with four different annotators (three law students and one employee in the municipal administration of city Jena, refer to Figure 2). The latter was refined during the real annotation phase based on the comments and requirements of the new annotators, but not fundamentally changed. This resulted in three versions of the annotation guideline, each of which is used in a specific phase of the real annotation phase as described in Section 3.2.2.

## 5. GerPS-NER Dataset

### 5.1. Conversion Scripts

A commonly used data format for annotated data is *IOB*, meaning Inner-Outer-Begin. It consists of a file where each line contains a token and the assigned label separated by whitespace. Here, we use *IOB2*, which was defined in the CoNLL-2002 shared task [7]. To indicate the boundaries of an annotation, it will start with *B* to indicate the start of a label, and with *I* to show that this token is part of an annotation but not the beginning. *IOB1* will only use *B-* before a label if immediately before the label another label is given. Tokens without an annotation are followed by *O*.

The intermediate corpora were generated using scripts that transforms the files in WebAnno TSV 3.3 format<sup>31</sup> exported from INCEpTION [19] into *IOB2*-format needed by the models for training and testing. As the projects in the real annotation phase differ in the style of annotation, for the conversion of each project, an individual script was necessary<sup>32</sup>. Though more information was collected in the annotation phases, we consider only the biggest annotated entities while generating the *IOB2* file, in case of overlapping annotations because there is not support for those in *IOB2*-format. The overlap was allowed starting from phase “Project 0.1” in the real annotation

<sup>27</sup>[1] in folder `annotation_guidelines/Annotation_Guide_V0.md`

<sup>28</sup>[1] in folder `annotation_guidelines/Annotation_Guide_V0.1.md`

<sup>29</sup>[1] in folder `annotation_guidelines/Annotation_Guide_V0.2.md`

<sup>30</sup>[1] in folder `annotation_guidelines/30-03-2023_Annotation_Guide_V1.pdf`

<sup>31</sup>Refer to [https://inception-project.github.io/releases/31.3/docs/user-guide.html#sect\\_webannotsv](https://inception-project.github.io/releases/31.3/docs/user-guide.html#sect_webannotsv) for more information

<sup>32</sup>[1] in folder `code/conversion_to_iob`

**Table 2**

Distribution of annotations per category using total (#) and fraction of overall annotations (%)

Category	#	%
Aktion	7776	26.1
Signalwort	4818	16.2
Handlungsgrundlage	3369	11.3
Bedingung	3231	10.8
Dokument	2925	9.8
Ergebnisempfänger	2729	9.1
Hauptakteur	1590	5.3
Mitwirkender	1431	4.8
Frist	941	3.1
Datenfeld	891	2.9

phase. This will allow catching as much information as possible during the annotation, and testing the predictions of nested entities using machine learning models as a future work. In addition, we also do not consider negation by the *IOB2* file generation.

## 5.2. Final Dataset

GerPS-NER is the accumulation of the following intermediate corpora: the short span version from adjudication phase (10 documents), project 0.1 (20), project 1 (29), and project 2 (775). Therefore, it consists of 834 documents with 24,613 sentences with 495,303 tokens. Of the tokens, 120,517 (24.3%) were part of an annotation (total annotated tokens), with 29,701 annotations in total (total annotations) as one annotation can consist of multiple tokens. Table 2 shows the distribution of the annotations over the different categories.

## 6. Conclusion and Future Work

We present GerPS-NER, a corpus of annotated German law texts using the extended FIM - norm analysis categories consisting of 24,613 sentences with 495,303 tokens in total and 29,701 annotations (consisting of one or more annotated tokens) and extensively describe the annotation process and guidelines. In future work, we will extend the corpus with ongoing annotated documents and provide further evaluation of its content including metrics and applications when comparing different techniques for annotation legal texts, as we will use the GerPS-NER corpus to test the effectiveness of rulebased methods (e.g., [20]), fine-tuning of different machine learning models, and for prompting with large language models (LLMs) [21]. As a start, Bachinger et al. [22] used a first corpus version (as described in Section 3.2.1) to examine the usage of different prompt variations on the performance of NER with LLMs where they report micro F1-scores for the optimal scenario of 0.82 for LeoLM [23].

## Acknowledgments

The research projects Canarèno, simpLEX, and KollOM-Fit of the working group openDVA were funded by the federal digitization budget and the federal government's fund in the scope of OZG implementation<sup>33</sup> with the support of the Federal Ministry of the Interior and Community, FITKO<sup>34</sup>, and Thuringian Ministry of Finance in Germany<sup>35</sup>. We would like to thank all employees, project partners, and supporters of the openDVA working group, who could not be mentioned here by name, for their great support, helpful comments, discussions, and good cooperation.

## References

- [1] L. Feddoul, S. T. Bachinger, S. Apel, P. Karg, N. Klewer, D. Forshayt, R. Erd, M. Mauch, C. Lachenmaier, Gerps-ner: Dataset and code, 2024. doi:10.5281/zenodo.10822682.
- [2] <https://fimportal.de/>, Normenanalyse, <https://fimportal.de/glossar>, 2024. Accessed: 02.01.2024.
- [3] T. O. M. Group, Business process model and notation, <https://www.omg.org/spec/BPMN/2.0/About-BPMN>, 2010. Accessed: 02.01.2024.
- [4] M. Mauch, S. T. Bachinger, P. Bornheimer, S. Breidenbach, D. Ehrhardt, L. Feddoul, H. Legner, F. Löffler, F. Löffler, M. Raupach, S. Schindler, J. Schröder, B. König-Ries, From legal texts to digitized services for public administration, in: Language Models: Legal Parrots or more? Proceedings of the 27th International Legal Informatics Symposium IRIS 2024, 2024. URL: <https://easychair.org/publications/preprint/PsVv>.
- [5] Y. Zhao, J. Liu, Human-in-the-loop based named entity recognition, in: 2021 International Conference on Big Data Engineering and Education (BDEE), 2021, pp. 170–176. doi:10.1109/BDEE52938.2021.00037.
- [6] J. Li, A. Sun, J. Han, C. Li, A survey on deep learning for named entity recognition, IEEE Transactions on Knowledge and Data Engineering 34 (2022) 50–70. doi:10.1109/TKDE.2020.2981314.
- [7] E. F. Tjong Kim Sang, Introduction to the CoNLL-2002 Shared Task: Language-Independent Named Entity Recognition, in: COLING-02: The 6th Conference on Natural Language Learning 2002 (CoNLL-2002), 2002. URL: <https://aclanthology.org/W02-2024>.
- [8] E. Leitner, G. Rehm, J. Moreno-Schneider, Fine-grained Named Entity Recognition in Legal Documents, in: M. Acosta, P. Cudré-Mauroux, M. Maleshkova, T. Pellegrini, H. Sack, Y. Sure-Vetter (Eds.), Semantic Systems. The Power of AI and Knowledge Graphs. Proceedings of the 15th International Conference (SEMANTiCS 2019), number 11702 in Lecture Notes in Computer Science, Springer, Karlsruhe, Germany, 2019, pp. 272–287. 10/11 September 2019.
- [9] E. Leitner, G. Rehm, J. Moreno-Schneider, A dataset of German legal documents for named entity recognition, in: Proceedings of the Twelfth Language Resources and Evaluation

---

<sup>33</sup><https://www.bmi.bund.de/SharedDocs/pressemitteilungen/DE/2021/02/ozg-konjunkturmittelverteilung.html>

<sup>34</sup><https://www.fitko.de>

<sup>35</sup><https://finanzen.thueringen.de>

- Conference, European Language Resources Association, Marseille, France, 2020, pp. 4478–4485. URL: <https://aclanthology.org/2020.lrec-1.551>.
- [10] H. Darji, J. Mitrović, M. Granitzer, A dataset of german legal reference annotations (2023).
- [11] M. Wrzalik, D. Krechel, GerDaLIR: A German dataset for legal information retrieval, in: N. Aletras, I. Androutsopoulos, L. Barrett, C. Goanta, D. Preotiuc-Pietro (Eds.), Proceedings of the Natural Legal Language Processing Workshop 2021, Association for Computational Linguistics, Punta Cana, Dominican Republic, 2021, pp. 123–128. URL: <https://aclanthology.org/2021.nllp-1.13>. doi:10.18653/v1/2021.nllp-1.13.
- [12] J. Pustejovsky, A. Stubbs, Natural Language Annotation for Machine Learning, number v. 9, p. 878 in A Guide to corpus-building for applications, O’Reilly Media, Incorporated, 2013. URL: <https://books.google.de/books?id=QtzmqamXxx4C>.
- [13] L. Feddoul, M. Raupach, F. Löffler, S. Babalou, J. Hoyer, M. Mauch, B. König-Ries, On which legal regulations is a public service based? fostering transparency in public administration by using knowledge graphs, in: INFORMATIK 2023 - Designing Futures: Zukünfte gestalten, Gesellschaft für Informatik e.V., Bonn, 2023, pp. 1035–1040. doi:10.18420/inf2023\_115.
- [14] J. Hoyer, Die Erstellung einer Prozessontologie zur Modellierung von Verwaltungsprozessen, Bachelor’s thesis, Jena, 2023. URL: [https://www.db-thueringen.de/receive/dbt\\_mods\\_00057705](https://www.db-thueringen.de/receive/dbt_mods_00057705), bachelorarbeit, Friedrich-Schiller-Universität Jena, 2023.
- [15] J. Cohen, A coefficient of agreement for nominal scales, Educational and Psychological Measurement 20 (1960) 37–46. URL: <https://doi.org/10.1177/001316446002000104>. doi:10.1177/001316446002000104. arXiv:<https://doi.org/10.1177/001316446002000104>.
- [16] Kataloge - katalog download leistungen, <https://fimportal.de/kataloge>, 2023. Accessed: 05.01.2024.
- [17] S. Bachinger, C. Lachenmaier, L. Feddoul, Data-collection-fim-laws, 2024. URL: <https://doi.org/10.5281/zenodo.10450554>. doi:10.5281/zenodo.10450554.
- [18] S. Bachinger, C. Lachenmaier, L. Feddoul, Corpus-fim-laws, 2023. URL: <https://doi.org/10.5281/zenodo.7900297>. doi:10.5281/zenodo.7900297.
- [19] J.-C. Klie, M. Bugert, B. Boullosa, R. E. de Castilho, I. Gurevych, The inception platform: Machine-assisted and knowledge-oriented interactive annotation, in: Proceedings of the 27th International Conference on Computational Linguistics: System Demonstrations, Association for Computational Linguistics, 2018, pp. 5–9. URL: <http://tubiblio.ulb.tu-darmstadt.de/106270/>, event Title: The 27th International Conference on Computational Linguistics (COLING 2018).
- [20] T. Eftimov, B. Koroušić Seljak, P. Korošec, A rule-based named-entity recognition method for knowledge extraction of evidence-based dietary recommendations, PLOS ONE 12 (2017) 1–32. URL: <https://doi.org/10.1371/journal.pone.0179488>. doi:10.1371/journal.pone.0179488.
- [21] W. X. Zhao, K. Zhou, J. Li, T. Tang, X. Wang, Y. Hou, Y. Min, B. Zhang, J. Zhang, Z. Dong, Y. Du, C. Yang, Y. Chen, Z. Chen, J. Jiang, R. Ren, Y. Li, X. Tang, Z. Liu, P. Liu, J.-Y. Nie, J.-R. Wen, A Survey of Large Language Models, Technical Report, 2023. URL: <http://arxiv.org/abs/2303.18223>. doi:10.48550/arXiv.2303.18223, arXiv:2303.18223 [cs] type: article.

- [22] S. T. Bachinger, L. Feddoul, M. Mauch, B. König-Ries, Extracting Legal Norm Analysis Categories from German Law Texts with Large Language Models, in: 25th Annual International Conference on Digital Government Research (dg.o 2024), June 11–14, 2024, Taipei, Taiwan, 2024. doi:10.1145/3657054.3657277.
- [23] B. Plüster, LeoLM: Igniting German-language LLM Research, <https://laion.ai/blog/leo-lm/>, 2023.