

# Codebook v 1.0

The Czech Constitutional Court Dataset  
The CCC Dataset

Štěpán Paulík

## Datasets

---

metadata . . . . .	2
texts . . . . .	6
judges . . . . .	7
clerks . . . . .	9
compositions . . . . .	11
separate_opinions . . . . .	12
parties . . . . .	13
references . . . . .	14
subject_matter . . . . .	15
disputed_acts . . . . .	16
verdicts . . . . .	17

# metadata

decision-level overview-metadata

## Description

This dataset includes overview information on all 93826 decisions of the CCC spanning from its founding in 1993 to the end of 2023. There is always one observation per decision. The datasets includes information on judge rapporteur, the deciding body, date of the submission and date of decision, as well as nested all the remaining datasets. The information contained herein unlocks the potential to wrangle the data to fit anyone's ends, for example, to create time-series data on caseload of different internal bodies of the CCC and alike.

## Variables

<code>doc_id</code>	<b>string</b> Document ID representing a unique identifier for each decision. The <code>doc_id</code> identifier is used across all the decision-level datasets.
<code>case_id</code>	<b>string</b> A case ID representing a unique identifier for the legal case. Any legal case can contain more than one decision.
<code>case_nr</code>	<b>numeric</b> An integer specifying the number that a decision has within a case.
<code>popular_name</code>	<b>string</b> Popular name associated with the case as reported by the CCC in its database
<code>date_decision</code>	<b>date</b> Date of the decision in the format YYYY-MM-DD.

<code>year_decision</code>	<b>numeric</b> Year of the decision.
<code>date_publication</code>	<b>date</b> Date of the publication in the format YYYY-MM-DD.
<code>date_submission</code>	<b>date</b> Date of the submission in the format YYYY-MM-DD.
<code>type_decision</code>	<b>string</b> Type of decision can take up the possible values of Nález (a decision on merits), Usnesení (a decision on admissibility or a procedural decision), or Stanovisko pléna (a decision whose purpose is to unify the CCC's caselaw).
<code>type_proceedings</code>	<b>string</b> Type of legal proceedings according to the constitution and the act on the CCC. The procedures contain concrete constitutional complain review ("O ústavních stížnostech"), abstract review of law ("O zrušení zákonů a jiných právních předpisů"), review of international agreements ("O souladu mezinárodních smluv"), separation of powers conflicts ("Ve sporech o rozsah kompetencí státních orgánů a orgánů územní samosprávy"), election and representative conflicts ("V pochybnostech o ztrátě volitelnosti a o neslučitelnosti výkonu funkcí poslance nebo senátora", "O návrhu politické strany dle čl. 87 odst. 1 písm. j) Ústavy", "O opravném prostředku proti rozhodnutí ve věci ověření volby poslance nebo senátora"), and impeachment against president procedure ("O ústavní žalobě Senátu proti prezidentu republiky")
<code>importance</code>	<b>numeric</b> An integer signifying the importance of the decision as scraped from the CCC's own database. The manual for the database does not explain the variable. The minimal possible value is 1, whereas the maximum is 4.
<code>applicant</code>	<b>nested-list</b> Nested part of the <code>parties</code> dataset specifying who the applicant was.
<code>concerned_body</code>	<b>nested-list</b> Nested part of the <code>parties</code> dataset specifying what the body concerned was.
<code>judge_rapporteur_name</code>	<b>string</b> Full name of the judge rapporteur.

<code>judge_rapporteur_id</code>	<b>string</b> Judge ID representing a unique identifier for each judge. The identifier can be used to join with the <code>judges</code> dataset.
<code>disputed_act</code>	<b>nested-list</b> Nested part of the <code>disputed_acts</code> dataset.
<code>type_verdict</code>	<b>nested-list</b> Nested part of the <code>verdicts</code> dataset.
<code>grounds</code>	<b>string</b> Details on the grounds of the decision regarding whether a decision is based on merits, admissibility or whether it was a procedural decision. The resulting grounds are inferred from the verdicts a decision contains. If the verdicts a decision contains are only procedural, then the grounds in sum are also deemed procedural. If all the verdicts are on admissibility, then the grounds are also in sum deemed on admissibility. If the verdicts contain any of the verdicts on ground, then the grounds for the decision in sum are deemed as on merits.
<code>concerned_constitutional_acts</code>	<b>nested-list</b> Nested part of the <code>references</code> dataset.
<code>concerned_acts</code>	<b>nested-list</b> Nested part of the <code>references</code> dataset.
<code>separate_opinion</code>	<b>nested-list</b> Nested part of the <code>separate_opinions</code> dataset.
<code>subject_proceedings</code>	<b>nested-list</b> Nested part of the <code>subject_matter</code> dataset.
<code>field_register</code>	<b>nested-list</b> Nested part of the <code>subject_matter</code> dataset.
<code>note</code>	<b>string</b> Additional notes or comments about the case. Mostly a link to press release accompanying a decision.
<code>url_address</code>	<b>string</b> URL address linking to more details about the case.
<code>formation</code>	<b>string</b> The judicial formation handling the case (e.g., plenum or a three member chamber).
<code>length_proceeding</code>	<b>numeric</b> Length of the legal proceeding in days from the date of the submission until the date of decision.
<code>outcome</code>	<b>string</b> Outcome of the case (e.g., granted, rejected).
<code>composition</code>	<b>nested-list</b> Nested part of the <code>compositions</code> dataset.

`citations`

`nested-list` Nested part of the `references` dataset.

# texts

decision-level text-corpus data

## Description

---

This dataset includes full text corpus of all CCC decisions. This dataset unlocks the potential for various quantitative text analysis or machine learning applications.

## Variables

---

`doc_id`

**string** Document ID representing a unique identifier for each decision.

`text`

**string** Full text of a decision.

# judges

judge-level judges background-data

## Description

---

This dataset includes judge-level information on all judges that have served at the CCC. The dataset contains information on their background, such as gender, highest degree, or alma mater. All information has been partly automatically and partly manually coded based on publicly available information.

## Variables

---

<code>judge_id</code>	<b>string</b> A judge ID representing a unique identifier across all judge-level or judge-including datasets.
<code>judge_name</code>	<b>string</b> The full name of the judge.
<code>judge_name_lemmatized</code>	<b>string</b> Lemmatized form of the judge's name.
<code>judge_yob</code>	<b>numeric</b> The year of birth of the judge.
<code>judge_gender</code>	<b>string</b> The gender of the judge.
<code>judge_uni</code>	<b>string</b> The university where the judge obtained their degree.
<code>judge_degree:</code>	<b>string</b> The highest academic degree obtained by the judge. The degrees go in the following order (desc.) prof > doc > phd > judr > mgr.



<code>judge_term_start</code>	<b>date</b> The start date of judge's term.
<code>judge_term_end</code>	<b>date</b> The end date of judge's term.
<code>judge_term_court</code>	<b>string</b> The term roughly corresponding to each decade of the CCC.
<code>judge_term_president</code>	<b>string</b> The name of the president of republic that appointed a judge for their term.
<code>judge_reelection</code>	<b>dummy</b> A dummy variable indicating whether a justice ran for more than one term. Whether the reelection attempt was successful or not can be discerned from the fact that the successful candidates have more than one observation in the <code>judges</code> dataset.
<code>judge_initials</code>	<b>string</b> Initials of a judge. The main use are regex searches of the texts of the decisions for various purposes.

# clerks

clerk-level clerks background-data

## Description

This dataset includes judge-level information on all clerks that have served under a judge at the CCC. The dataset contains information on their background, such as gender, highest degree, or whether they studied abroad. All information has been partly automatically and partly manually coded based on publicly available information.

## Variables

<code>judge_id</code>	<code>string</code>	Judge ID of the judge under whom the clerk worked.
<code>judge_name</code>	<code>string</code>	Name of the judge under whom the clerk worked.
<code>clerk_name</code>	<code>string</code>	The full name of the clerk.
<code>clerk_ID</code>	<code>string</code>	A clerk ID representing a unique identifier.
<code>clerk_term_start</code>	<code>date</code>	The start date of clerk's term.
<code>clerk_term_end</code>	<code>date</code>	The end date of clerk's term.
<code>clerk_name_full</code>	<code>string</code>	The full name of the clerk, including their academic titles.
<code>clerk_gender</code>	<code>string</code>	The gender of the clerk.

**clerk\_degree**

**string** The highest academic degree obtained by the clerk. The degrees go in the following order (desc.) prof > doc > phd > judr > mgr.

**clerk\_abroad**

**dummy** A dummy variable indicating whether the clerk obtained one of their degree abroad. The data was obtained from the **clerk\_name\_full** under the assumption that the Czech legal degrees differ from the typical LLB and LLM degrees.

# compositions

decision-level composition data

## Description

---

This dataset contains information on which judges sat at which case. Each observation is a composite observation of a decision and a judge that took part in that decision. The data has been mined automatically from the text corpus using regular expressions. Because of a rather high degree of irregularity of the decisions from the first decade, the data is not entirely reliable for that period.

## Variables

---

<code>doc_id</code>	<code>string</code> Document ID representing a unique identifier for each decision.
<code>judge_name</code>	<code>string</code> The full name of the judge sitting at the case.
<code>judge_id</code>	<code>string</code> The judge ID of the judge sitting at the case.

# separate\_opinions

decision-judge-level separate-opinion data

## Description

This dataset contains information on which judges attached a separate opinion to a decision. Each observation is a composite observation of a decision and a judge that attached a separate opinion to that decision. The information on whether a judge attached a separate opinion or not comes from the CCC's own database, the remaining data has been to some extent mined automatically from the text corpus using regular expressions. Because of a rather high degree of irregularity of the decisions from the first decade, the data is not entirely reliable for that period.

## Variables

<code>doc_id</code>	<code>string</code> Document ID representing a unique identifier for each decision.
<code>dissenting_judge_name</code>	<code>string</code> The full name of the judge who attached a separate opinion under the majority decision.
<code>dissenting_judge_id</code>	<code>string</code> The judge ID of the judge who attached a separate opinion under the majority decision.
<code>dissenting_group</code>	<code>numeric</code> The CCC judges can elect to dissent either alone or together with other people. The <code>dissenting_group</code> variable captures this distinction. When the judges attached a separate opinion together, they fall into the same <code>dissenting_group</code> .

# parties

decision-party level parties data

## Description

---

This dataset contains the information on parties, both the applicant and the involved party, as well as additional information on the type of the party. Each observation is a decision-party composite.

## Variables

---

<code>doc_id</code>	<b>string</b> Document ID representing a unique identifier for each decision.
<code>party</code>	<b>string</b> The full name and specification of a party in any given court proceedings before the CCC as obtained from the CCC database.
<code>party_type</code>	<b>string</b> Specification of the side at which the party stood before the CCC. Either they were the applicant party or the concerned body. A concerned body is typically the body, whose decision or legal act is being reviewed by the CCC. However, the court proceedings before the CCC do not have a contradictory form of proceedings.
<code>party_kind</code>	<b>string</b> The nature of the party, i.e., whether they are a natural person, legal person, state authority, court, or, for example, state prosecution.

# references

decision-reference-level references acts data

## Description

---

This dataset contains the information on concerned legal acts, that is both ordinary statutes as well as constitutional acts, and references to CCC caselaw. The information on the concerned acts come from the CCC itself, whereas the references to the CCC caselaw have been automatically mined from the text corpus using regular expressions.

## Variables

---

<code>doc_id</code>	<b>string</b> Document ID representing a unique identifier for each decision.
<code>concerned_act</code>	<b>string</b> The col. number of the concerned act in the "number/year Sb." format with information on the specific paragraph whenever such an information was available. In the case of CCC caselaw, the concerned act is in the standard citation form of "formation. ÚS number/year", which can then further be linked to the <code>case_id</code> variable from the <code>decisions</code> dataset.
<code>concerned_act_type</code>	<b>string</b> The type of the concerned act. Either an ordinary act (statute), a constitutional act (the Constitution, the Czech Charter of Fundamental Rights and Freedoms), or a CCC caselaw reference.

# subject\_matter

decision-subject-matter level data

## Description

---

This dataset contains information on the subject matter of cases. Each observation is a decision and one subject matter composite. The data was created by merging two information from the CCC database, namely on the subject proceedings and the concerned area of constitutional law.

## Variables

---

<code>doc_id</code>	<b>string</b> Document ID representing a unique identifier for each decision.
<code>subject_matter</code>	<b>string</b> The subject matter of the case as obtained from the CCC database.
<code>source</code>	<b>string</b> The CCC database contains two entries: subject of proceedings (předmět řízení) or subject register (věcný rejstřík). Both entries are very similar, although subject register mostly concerns areas of constitutional law (which constitutional right such as freedom of speech was infringed), whereas the subject proceedings mostly refers to the actual subject matter of proceedings. For clarity, the information about the source is kept.



# disputed\_acts

decision-disputed-acts-level data

## Description

---

This dataset includes the disputed acts in any given decision. Each observation is a decision-disputed act composite. The data was obtained from the CCC database. Additionally, information on the type of the disputed act was automatically added.

## Variables

---

<code>doc_id</code>	<b>string</b> Document ID representing a unique identifier for each decision.
<code>disputed_act</code>	<b>string</b> The disputed act in question in full.
<code>disputed_act_type</code>	<b>string</b> The type of the disputed act. The possible values range from a <code>court_decision</code> , <code>administrative_decision</code> , <code>statute</code> , <code>municipal_statute</code> , to various government acts and legal acts specific to the Czech context.

# verdicts

decision-verdict-level data

## Description

This dataset includes information on the type of verdict the CCC reached in any given decision, i.e., whether the case was granted, rejected, rejected on inadmissibility, or whether it was a procedural verdict. Each observation is a decision-verdict composita. Each decision may contain more than one verdicts.

## Variables

`doc_id`

**string** Document ID representing a unique identifier for each decision.

`verdict_type`

**string** The type of the verdict (výrok). A verdict is the binding ruling of the CCC. In general, vyhověno equals to granted. Zamítnuto and odmítnuto are both rejections, with the former being on merits and the latter on admissibility. The inadmissibility can be based on many grounds roughly mirroring those of ECHR's admissibility criteria. Lastly, the procesní verdicts are procedural verdicts.

`verdict_ground`

**string** The grounds of each verdict. The possible values are merits, admissibility, and procedural. For a more detailed explanation, see the description of `verdict_type`.