

---

# BRAZILIAN JUSTICE PROCEEDINGS DATASET

---

## 1 Dataset

The Brazilian Justice Processes dataset contains finished and ongoing lawsuits from thousands of Brazilian courts. These courts are distributed among six distinct justice entities named *Justiça Federal*, *Justiça Estadual*, *Justiça Eleitoral*, *Justiça do Trabalho* and *Tribunais Superiores*.

The dataset contains public information provided in JSON format by Brazilian *National Justice Council* (CNJ) during a hackathon competition. CNJ, in its turn, extracted the data from the DATAJUD database, a unified system containing Brazilian lawsuit data. CNJ recommended caution when using the dataset since it contains only a sample of a much larger dataset.

The dataset was made available as multiple JSON files distributed in six folders. The folders are *justica\_eleitoral*, *justica\_militar*, *justica\_trabalho*, *justica\_federal*, *justica\_estadual* and *tribunais\_superiores*. Each folder represents a justice entity. In each folder, there are multiple subfolders and JSON files. Each JSON file lists lawsuit instances represented by their respective header and procedural acts (movements). Sensitive information such as related parties and documents are absent in the dataset.

### 1.1 Dataset Fields

Each folder of the dataset represents a justice realm. For each lawsuit, there is header information and the procedural movements taken by the process while being judged. Table 1 shows the most relevant fields. Each lawsuit may contain only some of the described fields. Header data is mainly inside the JSON field *dadosBasicos*. The procedural movements, in its turn, are placed inside the JSON field *movimento*. The *dadosBasicos.assunto* field contains all subjects related to the lawsuit. The lawsuit subject relates to the matters of law discussed in the case. A code identifies each subject. The field *dadosBasicos.assunto.codigoNacional* contains the subject code that is standard and used nationally-wide. In contrast, the field *dadosBasicos.assunto.assuntoLocal.codigoAssunto* refer to the code locally used by the court. When the *codigoNacional* is absent, one can usually map the local code to the national equivalent through the field *dadosBasicos.assunto.assuntoLocal.codigoPaiNacional*. However, this mapping may not be precise because the value of the standard code can be at a coarser level of granularity than the local code. The *dadosBasicos.classeProcessual* field contains the standard lawsuit type code. The lawsuit-type field refers to procedures adopted in the judicial realm to handle the process. It defines the type of process analysed in conjunction with the subjects. Consequently, they can correspond to trace-level attributes. Examples of lawsuit types are *Processo Militar* and *Processo Cível e do Trabalho*—finally, The *dadosBasicos.orgaoJulgador.codigoOrgao* contains the code of the court that handled the process.

The fields inside *movimento* variable in the JSON correspond to the lawsuit body and capture each change in the state of the process while being judged. Therefore, they can correspond to trace activities. The field *movimento.movimentoNacional.codigoNacional* contains a standard code that nationally identifies the movement. Correspondingly, the field *movimento.movimentoLocal.codigoMovimento* contains a code specific to the court responsible for the movement. Similarly to *dadosBasicos.assunto* field, one can retrieve the equivalent national code through the field *movimento.movimentoLocal.codigoPaiNacional*. However, as previously mentioned, it is essential to be cautious with this mapping because they can not correspond to the same level of granularity of the movement. The field *movimento.dataHora* captures the date and time when the procedural movement occurred.

Table 1: Description of fields in lawsuitdata

Field	Type	Description
<code>dadosBasicos.assunto</code>	array	List containing all subjects related to the lawsuit
<code>dadosBasicos.assunto.principal</code>	boolean	Reveals if the given subject is the main subject of the lawsuit
<code>dadosBasicos.assunto.codigoNacional</code>	int	Standard code for the given lawsuitsubject
<code>dadosBasicos.assunto.assuntoLocal.codigoAssunto</code>	int	Local court code for the given subject
<code>dadosBasicos.assunto.assuntoLocal.codigoPaiNacional</code>	int	Equivalent standard subject code for local subject code
<code>dadosBasicos.numero</code>	string	Number given to each lawsuit
<code>dadosBasicos.classeProcessual</code>	int	Standard code for the given lawsuitclass
<code>dadosBasicos.orgaoJulgador.nomeOrgao</code>	string	Name of the court responsible for the lawsuit
<code>dadosBasicos.orgaoJulgador.codigoOrgao</code>	int	Code of the court responsible for the lawsuit
<code>dadosBasicos.orgaoJulgador.instancia</code>	string	Judging instance of the process. It can be ORIG: original instance where the lawsuit has started, REV: revision instance of a process originated in another instance, ESP: special revision instance of a lawsuit, EXT: extraordinary revision instance of a lawsuit and ADM: management analysis instance.
<code>grau</code>	string	Identifies the level of jurisdiction where the lawsuit is found. Possible values are SUP: superior court, G2: second degree, G1: first degree, TR: appeal court, JE: special court, TRU: regional uniformization court, TNU: national uniformization court, TEU: state uniformization court, CJF: federal justice council and CSJT: superior council of labor justice.
<code>movimento</code>	int	List containing all movements related to the lawsuit
<code>movimento.movimentoLocal.codigoMovimento</code>	int	Local court code for the given lawsuit movement
<code>movimento.movimentoLocal.codigoPaiNacional</code>	int	Standard father code of the local movement code
<code>movimento.movimentoNacional.codigoNacional</code>	int	Standard code for the given lawsuit procedural movement
<code>movimento.dataHora</code>	int	Timestamp field for each lawsuit procedural movement

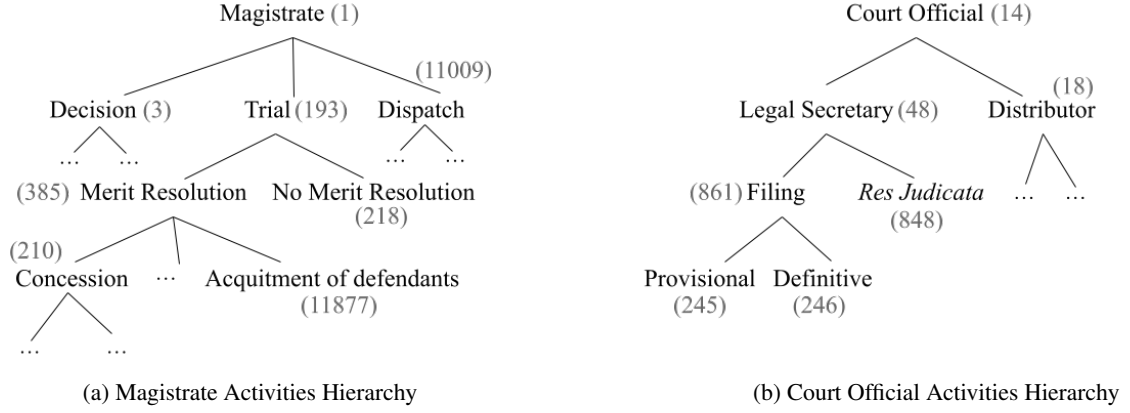
## 1.2 Hierarchical Attributes

The Brazilian Justice dataset contains some fields following a hierarchy. These fields are *dadosBasicos.assunto.codigoNacional*, *dadosBasicos.classeProcessual* and *movimento.movimentoNacional.codigoNacional* as described in Table 1. The hierarchy allows one to change the granularity of the attribute to a coarser level. Fig. 1 illustrates the two possible hierarchies for attribute *movimento.movimentoNacional.codigoNacional*. Suppose code 245 is a procedural movement of a particular lawsuit in the dataset. As the figure shows, this code will correspond to the movement *Provisional*. This movement can be mapped to *Filing* or *Legal Secretary* up to *Court Official*. Each higher level in the hierarchy contains a more generic description of the movement. The advantage of such a change is that it can make traces more homogeneous. This mapping can be critical because the dataset contains many unique traces. With this and some additional preprocessing, different lawsuits can become more comparable. The mapping from a code to a name of the attribute and the hierarchies are present in the files *df\_classes.csv*, *df\_assuntos.csv* and *df\_movimentos.csv*. They respectively correspond to fields *dadosBasicos.assunto.codigoNacional*, *dadosBasicos.classeProcessual* and *movimento.movimentoNacional.codigoNacional*. In all three files, the *breadscrum* field will have code sequences that define the attribute’s position in the tree. For instance, the code for movement *Provisional* will be 14.48.861.245. Each portion of the number will correspond to a level in the hierarchy down to the *Provisional* movement, as seen in Fig. 1b.

## 1.3 Creation of Event Log

A code for creating an example event log from a part of the dataset is available in Github<sup>1</sup>. In the code, it is also possible to apply additional preprocessing.

<sup>1</sup><https://github.com/lfvvercosa/brazilian-justice>

Figure 1: The two hierarchies of *Movement* attribute

#### 1.4 Additional Files

The files *Glossário-Datajud-Processos.docx* and *Informações Complementares CNJ Inova - Desafios e Dados.pdf* contain additional information, in Portuguese, about the dataset and the hackathon where it was made available.