

## Codebook for DAKSH High Court Database (2023)

1. Introduction
2. Overview of fields in DAKSH High Court Database
  - a. Bail (2023)
  - b. Writ (2022)
3. Cleaning and Analysis Protocols
4. Description of fields in datasets

**Data Source:** DAKSH has collected data through the eCourts Services mobile application, developed by NIC. Data pertaining to case status within ambit of High Courts is maintained by eCourt India Services. However, actual content relating to case status has been provided by respective High Courts.

**Data Availability Statement:** Datasets and visualisations provided on the High Court Data Portal are based on data collected through the eCourt India Services app. While all efforts have been made to ensure correctness of information, data availability gaps and other limitations may impact our findings.

The DAKSH database seeks to make data regarding the judiciary available in an accessible format for public engagement and advocacy. Any inadvertent errors are those of DAKSH alone. DAKSH may update, supplement, or remove material from these datasets at any time.

Datasets within the DAKSH High Court Database dataset are made available under the Creative Commons Attribution-Non Commercial 4.0 International license (CC BY-NC 4.0).

### **Preferred citation(s):**

DAKSH India (2022). DAKSH High Court Database: Writ Case dataset (v2). [url]

DAKSH India (2023). DAKSH High Court Database: Bail Case dataset. [url]

This is Version 2.0 of this Codebook. Further changes to this document and the data preparation process may be made at any time.

## Introduction

From 2016 onwards, DAKSH India has been involved in collating and visualising data from District & Sessions Courts in India as part of an effort to understand case volumes, pendency, and quality of judicial data.<sup>1</sup> This initiative, which grew into the **Rule of Law project**, developed a protocol for addressing data quality issues while maintaining the DAKSH database.

The existing procedure had to be adapted when applied to data gathered from various High Court across India; certain fields are populated differently at this level and the task of mapping and categorising case types and stages within cases takes on greater value. The key challenge is that High Courts follow different conventions when it comes to naming & classifying cases – which makes it difficult to harmonise judicial data gathered from across the country.

This document outlines some of the issues observed and processes followed while preparing High Court data for analysis, visualisation, and further research.

## Overview of fields in DAKSH High Court Database

### *Bail Case dataset*

Sl. No.	Field	Data type	By DAKSH <sup>2</sup>
1	CNR_NUMBER	Alphanumeric identifier	
2	CASE_NUMBER	Numeric identifier	
3	CASE_TYPE	Character (case attribute)	
4	CASETYPE_FULLFORM	Character (case attribute)	Y
5	CIVIL_CRIMINAL	Character (case attribute)	Y
6	SUB_CLASSIFICATION	Character (case attribute)	Y
7	COMBINED_CASE_NUMBER	Alphanumeric identifier	
8	COURT_NAME	Character (court attribute)	
9	COURT_NUMBER	Character (court attribute)	
10	NAME_OF_HIGH_COURT	Character (court attribute)	
11	CURRENT_STAGE	Character (case attribute)	
12	CURRENT_STATUS	Character (case attribute)	
13	DATE_FILED	Date	
14	DECISION_DATE	Date	
15	FILING_NUMBER	Character (case attribute)	
16	HEARING_COUNT	Number	
17	LAST_SYNC_TIME	Date	Y
18	NATURE_OF_DISPOSAL	Character (case attribute)	
19	NATURE_OF_DISPOSAL_OUTCOME	Character (case attribute)	Y
20	NJDG_JUDGE_NAME	Character (person)	
21	PENDING_DAYS	Number	Y
22	POLICE_STATION	Character (place)	

---

<sup>1</sup> DAKSH has already written about challenges of working with Judicial Data [here](#)

<sup>2</sup> “By DAKSH” is a field that was not scraped from eCourts but further validated, populated with the assistance of practicing lawyers, or calculated by DAKSH India from given date/numeric fields.

23	REGISTRATION_DATE	Date	
24	REGISTRATION_NUMBER	Character (case attribute)	
25	RESPONDENT	Character (person)	
26	UNDER_ACTS	Character (case attribute)	
27	UNDER_SECTIONS	Character (case attribute)	
28	YEAR	Year	
29	NATURE_OF_DISPOSAL_BINARY	Character (case attribute)	Y
30	DISPOSAL_YEAR	Year	Y
31	DISPOSAL_DAYS...1	Number	Y
32	Mapped_Bail	Character (case attribute)	Y

*Writ Case dataset*

Sl. No.	Field	Data type	By DAKSH
1	CNR_NUMBER	Alphanumeric identifier	
2	CASE_NUMBER	Numeric identifier	
3	CASE_TYPE	Character (case attribute)	
4	CASETYPE_FULLFORM	Character (case attribute)	Y
5	CIVIL_CRIMINAL	Character (case attribute)	Y
6	SUB_CLASSIFICATION	Character (case attribute)	Y
7	COMBINED_CASE_NUMBER	Alphanumeric identifier	
8	COURT_NAME	Character (court attribute)	
9	COURT_NUMBER	Character (court attribute)	
10	NAME_OF_HIGH_COURT	Character (court attribute)	
11	CURRENT_STAGE	Character (case attribute)	
12	CURRENT_STATUS	Character (case attribute)	
13	DATE_FILED	Date	
14	DECISION_DATE	Date	
15	FILING_NUMBER	Character (case attribute)	
16	HEARING_COUNT	Number	
17	LAST_SYNC_TIME	Date	Y
18	NATURE_OF_DISPOSAL	Character (case attribute)	
19	NATURE_OF_DISPOSAL_BINARY	Character (case attribute)	Y
20	NJDG_JUDGE_NAME	Character (person)	
21	PENDING_DAYS	Number	Y
22	POLICE_STATION	Character (place)	
23	REGISTRATION_DATE	Date	
24	REGISTRATION_NUMBER	Character (case attribute)	
25	RESPONDENT	Character (person)	
26	UNDER_ACTS	Character (case attribute)	
27	UNDER_SECTIONS	Character (case attribute)	
28	YEAR	Year	
29	DISPOSALTIME_ADJ	Number	Y
30	DISPOSAL_YEAR	Year	Y
31	CASE_CATEGORY	Character (case attribute)	Y
32	DISPOSAL_PATTERN	Character (case attribute)	Y

## Cleaning and Analysis Protocols

Data collection was done in January and February 2021 from 23 High Courts (out of 25) across India. It was then repeated during the months of June to December 2022.

In our scraping process, we obtained 87 fields at the case level. However, many of these are not consistently filled during data entry and contain numerous null values.

Out of these, 22 fields are selected for our case dataset – these fields are further used as the basis to expand, manually map, and calculate details for an additional 10 fields (identified as **By DAKSH** in the above tables).

As part of the cleaning process:

- I. Cases were de-duplicated & filtered; V2.0 is predominantly interested in cases filed between 2010 and 2021.  
  
Digitisation of court records increased significantly post-2010 under the eCourts Mission. Selecting the cut-off disposal year of 2021 has two benefits – (i) it allows us to observe cases filed during the first year of the COVID-19 pandemic (~Mar 22), yet (ii) it does not allow scraping dates spread across 2022 to provide an incomplete picture of cases filed that calendar year.
- II. Discrepancies between fields that should concur (i.e. Date of filing & Year) were adjusted in accordance with the field that provided greater detail.
- III. Researchers at DAKSH then applied their knowledge of the court systems to map relevant fields. For High Court data, this included the nature of case (Civil, Criminal or Writ), level of appeal, the expansion of the abbreviated Case Type, and nature of disposal.
- IV. Calculations were generated for Disposal days and Pending days and formulas used were discussed with peer research organisations.
- V. Final checks were done to ensure no essential fields were empty.
- VI. In the interest of protecting personal data of petitioners are advocates, the ‘Petitioner’ field has been removed from the dataset being released.

We have maintained a version with these fields included. Please reach out to DAKSH at [info@dakshindia.org](mailto:info@dakshindia.org) if you are interested in accessing this data for research.

In V2.0 of this project, the goal is to put out a usable dataset of High Court judicial data and collaborate with like-minded researchers and organisations to determine what insights could be reached at a system-level. We hope to collaborate to identify easy-to-implement modifications in data structuring that could better enable rigorous pan-India analysis of judicial workload.

## Description of fields in High Courts datasets

### *For Bail Case dataset –*

- CNR\_NUMBER is a 16-digit alphanumeric code assigned to each case that is unique across India.
  - Example: UPHC010933632010; the first four digits indicate which High Court the case originated in, and the last four digits indicate the year of filing
- CASE\_NUMBER is a number assigned to a case by each High Court Registry during the process of filing, scrutiny, and registration of cases.
- CASE\_TYPE is an abbreviation maintained by each High Court Registry that may indicate either substantive or procedural nature of each case.
- CASETYPE\_FULLFORM was generated by DAKSH and is the expansion of the abbreviated Case Type form. This information was manually collected by researchers at DAKSH for each High Court as an indexed list and populated across the entire dataset.
  - Example: BA in Jammu and Kashmir High Court stands for BAIL APPLICATION;
- CIVIL\_CRIMINAL was manually mapped by lawyers at DAKSH and may either be CIVIL, CRIMINAL, WRIT or OTHER. For the entire bail dataset, it has been populated as CRIMINAL.
- SUB\_CLASSIFICATION was manually mapped by lawyers at DAKSH and is an indication of the level of appeal, where that may be extrapolated from the case type.
  - Example: Regular First Appeal would be identified as a First Appeal
- COMBINED\_CASE\_NUMBER is the Case Number or Registration Number and is commonly used within each High Court to search for [case status details](#) within eCourts.
  - Example: BAILC-34022-2012; it is composed of three components: the abbreviated Case Type, the Case Number and the Year the case was registered.
- COURT\_NAME is the name of the High Court Bench (this may be location or purpose) before which the case appeared.
  - Example(s): High Court Bench at Jaipur; Circuit Bench At Jalpaiguri; Principal Bench, Appellate Side
- COURT\_NUMBER is a code assigned by eCourts to each Bench.
- NAME\_OF\_HIGH\_COURT is the name of the High Court before which the case appeared.
  - Example(s): High Court of Bombay, Orissa High Court, The Gauhati High Court
- CURRENT\_STAGE is the purpose or stage of the last recorded hearing held within the case. This is not standardized across benches or High Courts.
  - Example(s): Admission, For Motion, For Orders, For Withdrawal, Adjourned
- CURRENT\_STATUS may either be “Disposed” or “Pending” and indicates whether the case has been concluded or is active / ongoing before the court.
- DATE\_FILED is the date on which the lawyer representing the petition filed the case with the High Court Registry.
- DECISION\_DATE (for Disposed cases) is the date on which the final judgment was officially given.
- FILING\_NUMBER is a number assigned to a case by each High Court Registry during the process of filing.

- HEARING\_COUNT is the number of Hearings that have been held within the case (as per eCourts records).
- LAST\_SYNC\_TIME is the date on which DAKSH scraped the case record from eCourts.
- NATURE\_OF\_DISPOSAL (for Disposed cases) is the manner in which the case was decided. It contains two elements which we have further separated into NATURE\_OF\_DISPOSAL\_BINARY, which may either be “Contested” or “Uncontested” and NATURE\_OF\_DISPOSAL\_OUTCOME, which is the latter half of the recorded field.
  - Example: Contested-ALLOWED
- NJDG\_JUDGE\_NAME is generically identified at the case level; DAKSH’s hearing level data records the names of which single judge or division bench judges presided over each hearing within the case.
  - Example(s): In case dataset, it may be Justice, Judge, Chief Justice, etc
- PENDING\_DAYS was calculated by DAKSH as the number of calendar days that had passed from the DATE\_FILED to the LAST\_SYNC\_TIME; for a pending case, it is how many days a case has been active as on the scraping date.
- POLICE\_STATION is not standardised across the dataset, but indicates in which Police Station a criminal matter originated i.e. was first investigated and charge-sheeted.
  - Example(s): Kotwali, Mahila Thana
- REGISTRATION\_DATE is the date on which the filed petition passed scrutiny of the High Court Registry and was registered.
- REGISTRATION\_NUMBER is a number assigned to a case by each High Court Registry during the process of registration.
- RESPONDENT is the person, office or institution that is named in the petition and obligated to respond to the relief/remedy requested. For the bail dataset, it is generally either the State Government or the Public Prosecutor’s office.
- UNDER\_ACTS is the legislation under which the case was filed. This may be a procedural or a substantive legislation.
- UNDER\_SECTIONS mentioned the sections within identified legislation that are connected with the case.
- YEAR is the year in which the case was filed.
- DISPOSAL\_YEAR was generated by DAKSH and is the year in which the case was disposed.
- DISPOSAL\_DAYS...1 was calculated by DAKSH as the number of calendar days that had passed from the DATE\_FILED to the DECISION\_DATE (both days inclusive); for a disposed case, it is how many days a case took to be disposed.
- Mapped\_Bail was manually mapped by lawyers at DAKSH and is either “Regular Bail”, “Anticipatory Bail” or “Cancellation”.