

Introducing Corpora of Judgments for the German Federal Court of Justice (Bundesgerichtshof)

Criminal Matters (1950-1999, 2020–2024) and
Civil Matters (2000-2024)

Seán Fobbe and Tilko Swalve

Ludwig-Maximilians-Universität München

Universiteit Leiden

19 June 2025

German Federal Court of Justice Datasets

Overview

The **German Federal Court of Justice (BGH)** is the highest court of appeal in civil and criminal matters in Germany.

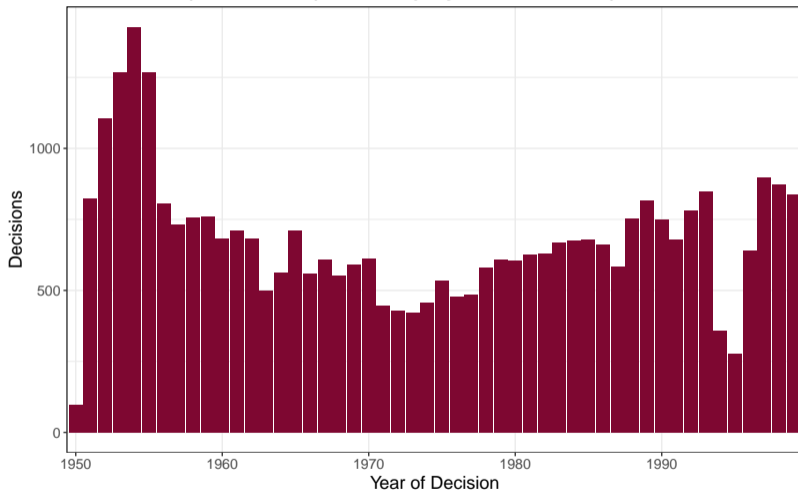
Dataset	CE-BGH	BGH-Strafsachen-20Jhd
Version	2025-04-07	1.0.0
Years	2000–2025	1950–1999
Decisions	77,892	36,316
Law	Civil and Criminal	Criminal
Variables	36 variables	31 variables
License	CC0 Public Domain	CC0 Public Domain

German Federal Court of Justice Datasets

Data Sources

- Original file format is PDF
- Decisions after 2000 available via official database
- Decisions 1950–1999 communicated from court via optical media to authors
- Sometimes poor scan quality for older decisions
- Limited metadata for older decisions

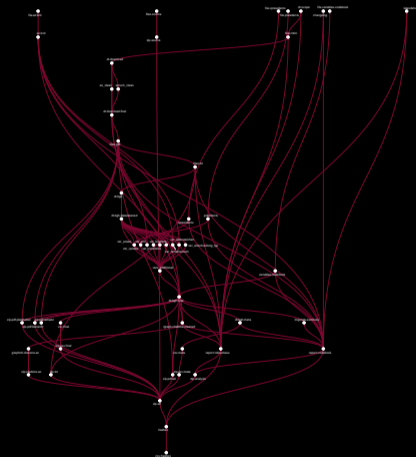
BGH Criminal | Version 1.0.0 | Decisions per year in the dataset | 2419 NA values



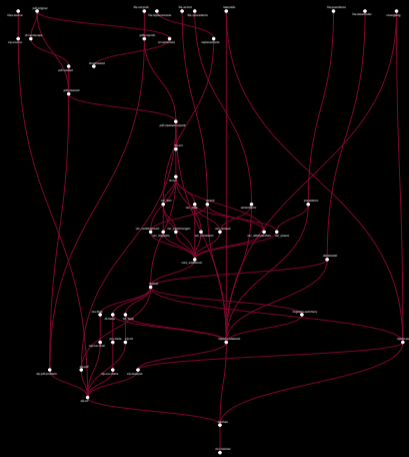
Legal Data Engineering

- End-to-end data pipeline in R and {targets} framework (Landau 2021)
- Recomputes components only if input, output or code have changed
- Fault-tolerant, resumable and auto-updating
- Automated testing to ensure syntactic correctness and plausibility of data
- Computational reproducibility with Rocker images for Docker (Nüst et al. 2020)
- Published Open Source (GPL-3)

End-to-End Data Pipelines



CE-BGH



BGH Criminal

Traditional Research

- **Original PDF files** with filename metadata for all decisions
- Specialized PDF variants for decisions from 2000
 - 1 Decisions with key legal rationales (*Leitsatzentscheidungen*)
 - 2 Named decisions (e.g. 'Trabrennbahn')
- **TXT files** for all decisions (born digital from 2000, Tesseract OCR for decisions 1950–1999)
- **Publication quality diagrams** for teaching and research

36 / 31 variables in 5 groups in CSV format

- 1 **[ID]** Date/year of decision, docket number, ECLI
- 2 **[Texts]** Full decision texts in German, basic length benchmarks
- 3 **[Context information]** Senate, name of decision, importance level (*Leitsatz*, BGHZ, BGHR etc), type of proceedings, comments
- 4 **[Source]** URLs to original PDF files
- 5 **[Meta]** Digital object identifiers and version information

Citation Network Data (Beta Phase)

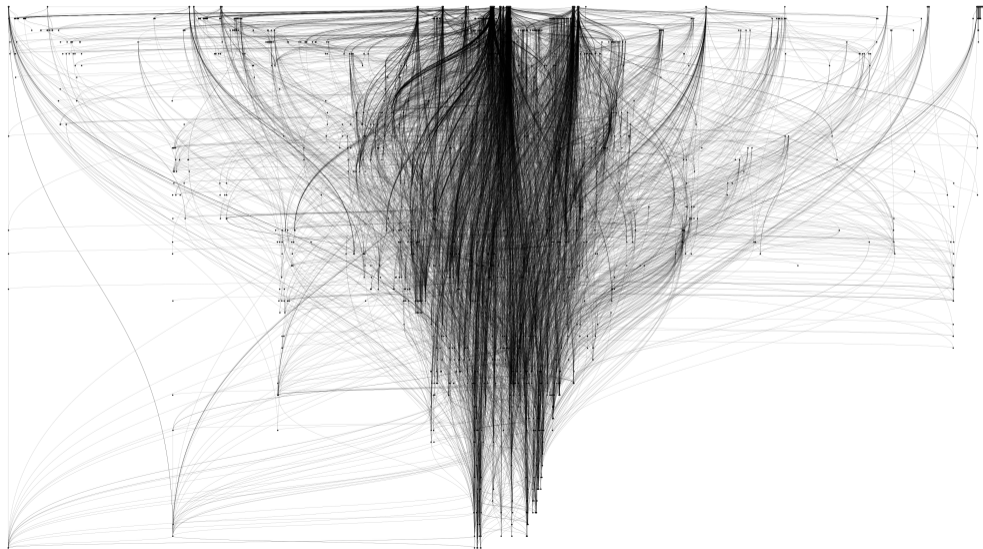
- Citations extracted from decision texts rendered 2000 and later
- 103,000 nodes (docket numbers and decisions from official collections)
- 612,000 individual citations, 456,000 weighted edges
- Intended as an easy-to-use source of citation data in GraphML format for exploration, prototyping and educational purposes

Example

The 'Trabrennbahn' decision can be cited as 'II ZR 158/06' (docket number) or 'BGHZ 178, 192' (official collection).

Internal Citation Network of the 11th BGH Senate

Docket Numbers and Official Collection of Decisions (BGHZ)



Download

German Federal Court of Justice (BGH) Datasets

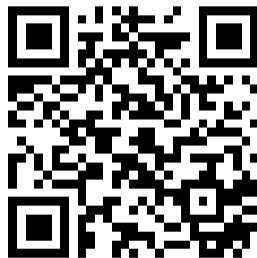
CE-BGH

2000–2025



BGH Criminal

1950–1999



www.seanfobbe.com/data