

The Integrated Database: A Research Guide

The Federal Judicial Center (FJC), under a working arrangement with the Administrative Office of the U.S. Courts (AOUSC), provides public access to its Integrated Data Base (IDB) through its webpage.¹ The IDB contains data on civil case and criminal defendant filings and terminations in the district courts, along with bankruptcy court and **appellate court case information**. This guide provides information for those seeking to use the data provided in the IDB, including how the data are collected, the time periods and data elements available, where to find the descriptions of the codes used, some basic usage tips, and notes on the data's limitations.

Collection Process

The FJC receives **quarterly updates of the case-related data** that are routinely reported by the courts to the AOUSC and published in the Judicial Business Reports. The FJC then post-processes the data, consistent with the policies of the Judicial Conference of the United States governing access to these data, into a unified longitudinal database, the IDB. The post-processing of the data takes several forms. First, data values that are out of range for the variable are recoded as missing. Second, some information is redacted—for example, the names of criminal defendants in criminal and appellate files. Relatedly, information on the judge or judges presiding over the case is redacted pursuant to Judicial Conference policy.² Third, the IDB integrates three types of case records: filings, pendings, or terminations. New cases are filings records. Cases that were filed previously but not yet terminated are pending records. Cases that were previously filed, or filed and terminated in the same quarter, are terminations records. Each quarterly update reflects the current status of the case records, including new case-related information (such as conversion of a bankruptcy from a Chapter 11 to a Chapter 7). These quarterly updates have implications for how best to use the IDB, discussed in more detail below.

Time Periods Available

The data available through the IDB vary in terms of the number of years available. Civil and criminal data are available back to 1970. **Data from the courts of appeals are available back to 1971**. Bankruptcy data are available back to 2008. Because the data-collection process has changed over time, and to allow for easier access to the data, the civil, criminal and appellate databases are divided into two files based on time periods. As detailed in the codebooks, and noted below, the lengthy duration of the available data creates some challenges for analyzing the

A previous version of this guide was prepared by members of the FJC Research Division as a white paper for the 2019 Liman Colloquium at Yale Law School.

1. All versions of the IDB can be accessed at <https://www.fjc.gov/research/idb>. The webpage provides the opportunity to download entire datasets or explore the data through an interactive view.

2. In March 1995, the Judicial Conference (JCUS) announced a policy “prohibiting the disclosure (except to the extent required by law) of judge-identifying information from statistical databases” (JCUS-MAR 95, pp. 21–22). The policy was reaffirmed by the JCUS at a meeting in September 1995 and again in March 2003.

data. The codebooks note the changes in the data-collection processes, including codes that have changed, that may require researchers to recode specific variables if they want to combine data files.

The courts report data for the statistical year, most notably in the Judicial Business Reports, and the information on the reporting year is provided in the IDB. Researchers may want to match their results to those reports, in which case they must be aware of changes to the statistical year over time. Up to 1992, the statistical year went from July through June (e.g., statistical year 1990 covered the period July 1, 1989, through June 30, 1990). In 1992, the statistical reporting period was changed to conform to the federal government's standard fiscal year, October through September (e.g., fiscal year 1993 covered the period October 1, 1992, through September 30, 1993). All of the previous data files in the IDB conform to the old statistical year (SY70–SY91). The 1992 files cover a 15-month time span (July 1, 1991, through September 30, 1992) to accommodate this conversion period. The 1993 file and all subsequent files conform to the new fiscal year (October 1 through September 30).

Basic Data Elements

The basic data elements available on the cases included in each version of the IDB are essentially the same.³ The circuit and district in which the case was filed, the office code, and the docket number combine to create a unique case identifier. Dates of filing and termination (if applicable) are available on each case, as is the type of termination.

Apart from these common fields, the information in each IDB file includes the case-level information relevant to each area of litigation. In the civil files, information on nature of suit, jurisdiction, origin codes, the names of plaintiffs and defendants, class action allegation, the procedural progress of the case at termination, and the nature and amounts of judgment are all included in the files.

The files for criminal litigation are structured somewhat differently to reflect the differences between civil and criminal litigation. Criminal cases often include multiple defendants, so the information provided in the criminal IDB is at the defendant (not case) level. The defendant number needs to be part of the unique identifier created for each observation. Included in the criminal IDB are the offense codes associated with the top five most severe counts, both at filing and termination, the number of the defendant in the case, the date of sentencing (if any), and the terms of sentencing for each offense code (if any), including separate fields for prison, probation, and supervised release. Aggregating the unique identifier across years will show how many defendants had data reported for that case.

The files for appellate data are a hybrid of the types of fields collected in both the civil and criminal data. The database contains information on the lower court proceedings as well as the appellate record, including which side appealed and the agency of the U.S. government involved

3. Only cases before district, appellate, or bankruptcy judges are reported in the IDB.

(if any). For cases originating in the courts of appeals, such as second or successive capital habeas petitions, the outcome is reported as granted or denied. For outcomes of cases from the lower courts, a combination of fields (disposition plus outcome or procedural termination or method) indicate the appellate court's decision.

Bankruptcy data involve substantially more variables than the other databases simply due to the nature of what is collected through the court filings. In addition to the petition, a schedule of assets and liabilities is submitted, providing information on the nature and amount of debt, as well as the availability of assets. Much of this information is included in the case record for the bankruptcy proceeding. Additionally, because bankruptcies can be moved from one Chapter to another, information about the original and current filing chapter, as well as the chapter at termination, is provided.

Codebooks

In cooperation with the Judiciary Data and Analysis Office of the AOUSC, the FJC has created and maintains codebooks that detail changes in the case data that have occurred since the creation of the IDB. These changes include details about new data fields that may have been added—for example, due to new legislation expanding the jurisdiction of the federal courts—or data elements that have been deleted and no longer routinely reported by the courts to the AOUSC. Additionally, the codebooks provide information on the source of information provided in the cases, and descriptions for the alphanumeric codes used in the databases themselves.

Guidelines for Use

While potential uses for IDB data are limited only by the researcher's imagination, there are some basic guidelines to be aware of when using these databases. First is how the data-collection process affects the information. In each quarterly update, the case records are replaced according to whether the case falls into the filed, pending, or terminated cohorts described above. In practice, this means that case information is potentially overwritten with each update if the record has been altered by the court. This may lead to some loss of information. For example, if a superseding indictment is filed, and the offenses involved change, the information on the first indictment may be lost. Similarly, if a replacement schedule of assets and liabilities is filed in a bankruptcy proceeding, the information will be replaced. Ultimately, the updating of records most accurately captures the specifics of the case as of the most recent snapshot. It does, however, inhibit a researcher's ability to know the full detail of the case record. For example, if a defendant was ever charged with a specific offense code, or if a litigant ever had counsel or was pro se for the entire duration of their case. If researchers are continually studying the federal courts, it would be

best to download each quarterly update and compare across records to have the most complete information on each case.⁴

A second issue for researchers to consider is how data quality may affect results. While many of the fields, especially with respect to those values in the published Judicial Business Reports, are collected with data quality control measures in place, some are not. A process of error correction and detection does not cover all the variables in the IDB, **though both the FJC and the AO make every effort to ensure the accuracy of the data.** Nonetheless, there may be some problems with specific fields that are not routinely reported.

The two issues with respect to data collection noted above are more likely to affect specific fields related to under-served populations. These fields include information regarding pro se litigants, *in forma pauperis* (IFP) status, and class action allegations.⁵ All three fields may change over the life of the case, making the issues with record replacement more pronounced in these fields, and all three are without data quality control checks, making it more likely that there are errors in reporting at any point across the duration of the case.

Conclusions

With the publicly available information on federal cases in district, bankruptcy, and appellate courts, researchers have an opportunity to explore patterns in federal litigation. The data available cover decades of cases and controversies, and the dozens of variables provided enable researchers to start exploring patterns immediately after download. Moreover, the case information available through the IDB gives researchers an accurate and complete account of case records from which they can sample or augment data available based on the needs of their research agenda or audience.

4. One caveat on the issue of updating records: if the status of a civil case changes within the same 12-month reporting period, the record is updated. If, however, the status of a civil case changes after the 12-month reporting period, the record will not appear. After the end of the Fiscal Year, records are not updated.

5. There are some business rules applied to the class action field. For example, the variable for terminating as a class action can't be coded as denied or granted if the filing class action variable isn't coded as yes. However, there are no quality-control checks on whether or not class allegations were made in the filing.