
The Federal Court Data Base: New research opportunities

by Terence Dungworth
The Rand Corporation
1700 Main Street
Santa Monica, ca 90406-2138

1. Introduction

The Research Division of the Federal Judicial Center has funded the standardization of the District and Circuit Court case records maintained by the Administrative Office of the U.S. Courts. The period covered is FY71 (fiscal year 1971) through FY85.

The product — referred to hereafter as the Federal Court Data Base (FCDB) — was placed in the public domain in the spring of 1986. The intent of the Federal Judicial Center is to update the FCDB at the end of each fiscal year, beginning with FY86. Sections I-V of this document summarize the standardization process and present an overview of the structure and content of the data base. For information concerning further documentation and tape availability, please contact the author.

II. The Reporting System of the Federal District and Circuit Courts

At the end of FY85, the Federal District Court system consisted of 95 districts, organized in 11 geographically structured circuits, plus a single circuit for Washington, D.C. Districts may have one or more offices, with the result that in some districts, cases may be filed and heard in different locations. A list of Circuits and Districts is presented in Table II-1.

All districts process the criminal and civil cases that fall within the jurisdiction of the federal court system. Circuit courts handle appeals from district court dispositions, and some original proceedings.

Certain information on every case filed is reported by each district and circuit to the Administrative Office of the US Courts (AO hereafter) in Washington, D.C. From this is created a central data base containing information on every federal case filed, terminated and appealed in the United States. The reported information on each case is compiled by the AO into a single case record, which, since about 1970, is maintained on magnetic tapes in machine readable format.

Records are grouped by general type — civil, criminal or appeals. Appeals may be from civil or criminal adjudications. At the end of each fiscal year, the AO produces three master tapes containing the records for cases terminated during that year, pending at year-end. Archive tapes are produced for the terminations, and these records are removed from the master tapes for the following year.

The AO also maintains records on cases filed in each fiscal year. These are not incorporated directly into the FCDB, but a comprehensive data base of filings can be constructed from a combination of FCDB terminations and pending files, provided the year(s) for which filings are needed is not earlier than the first year covered by the FCDB (1971).

TABLE 1

FEDERAL CIRCUIT AND DISTRICT COURTS

0090 District of Columbia	
First Circuit	Fifth Circuit
0100 Maine	053L Louisiana, Eastern
0101 Massachusetts	053N Louisiana, Middle
0102 New Hampshire	0536 Louisiana, Western
0103 Rhode Island	0537 Mississippi, Northern
0104 Puerto Rico	0538 Mississippi, Southern
	0539 Texas, Northern
	0540 Texas, Eastern
Second Circuit	0541 Texas, Southern

- 0205 Connecticut
- 0206 New York, Northern
- 0207 New York, Eastern
- 0208 New York, Southern
- 0209 New York, Western
- 0210 Vermont

Tenth Circuit

- 0311 Delaware
- 0312 New Jersey
- 0313 Pennsylvania, Eastern
- 0314 Pennsylvania, Middle
- 0315 Pennsylvania, Western
- 0391 Virgin Islands

Fourth Circuit

- 0416 Maryland
- 0417 North Carol., East.
- 0418 North Carol., Middle
- 0419 North Carol., West.
- 0420 South Carolina
- 0421 Virginia, Eastern
- 0422 Virginia, Western
- 0423 West Virginia, Northern
- 0424 West Virginia, Southern

Eighth Circuit

- 0860 Arkansas, Eastern
- 0861 Arkansas, Western
- 0862 Iowa, Northern
- 0863 Iowa, Southern
- 0864 Minnesota
- 0865 Missouri, Eastern
- 0866 Missouri, Western
- 0867 Nebraska
- 0868 North Dakota
- 0869 South Dakota

Ninth Circuit

- 097X Alaska

0542 Texas, Western

Sixth Circuit

- 0643 Kentucky, Eastern
- 0644 Kentucky, Western
- 0645 Michigan, Eastern
- 0646 Michigan, Western
- 0647 Ohio, Northern
- 0648 Ohio, Southern
- 0649 Tennessee, Eastern
- 0650 Tennessee, Middle
- 0651 Tennessee, Western

Seventh Circuit

- 0752 Illinois, Northern
- 0753 Illinois, Central
- 0754 Illinois, Southern
- 0755 Indiana, Northern
- 0756 Indiana, Southern
- 0757 Wisconsin, Eastern
- 0758 Wisconsin, Western

Tenth Circuit

- 1082 Colorado
- 1083 Kansas
- 1084 New Mexico
- 1085 Oklahoma, Northern
- 1086 Oklahoma, Southern
- 1087 Oklahoma, Western
- 1088 Utah
- 1089 Wyoming

Eleventh Circuit

- 1126 Alabama, Northern
- 1127 Alabama, Middle
- 1128 Alabama, Southern

0970 Arizona	1129 Florida, Northern
0971 California, Northern	113A Florida, Middle
0972 California, Eastern	113C Florida, Southern
0973 California, Central	113E Georgia, Northern
0974 California, Southern	113G Georgia, Middle
0975 Hawaii	113J Georgia, Southern
0976 Idaho	
0977 Montana	
0978 Nevada	
0979 Oregon	
0980 Washington, Eastern	
0981 Washington, Western	
0982 Guam	

III. The Federal Court Data Base Project

The Federal Court Data Base Project (FCDBP hereafter) was funded by the Research Division of the Federal Judicial Center. Its primary objectives were to convert existing machine-readable data into a standardized format.

The need for standardization arose because changes had been made over the years both in variable values that were legitimate and in the formats used to maintain records. In addition, in earlier years, range checking and other validation techniques had not been systematically applied. In consequence, invalid codes could be present in the data for any given year, and inter-year consistency of data structure and content had not been established. This made research time consuming and expensive, and seriously inhibited the utilization of an extremely valuable data resource.

Work on the data was done in two stages. First, the content of all fields for each year was examined, evaluated and cleaned; second, the cleaned data were put into a common format for all years.

The cleaning process was performed on a record-by-record basis. First, values were checked for valid range and format. Frequency distributions of all non-continuous variables were then produced and compared with the list of valid codes for that year. The latter were obtained from AO codebooks and data reporting forms used by the District and Circuit courts. Continuous variables — such as docket numbers, dollar figures, number of months given in a criminal sentence — were evaluated by other approaches.

Invalid codes which were discovered were examined to determine whether or not they could be converted. Frequently this was possible. A simple illustration is a data field that should have been in MMY format but was reported in YYMM format. This situation was corrected by switching the

YYMM fields to MMY Y. Another example is docket numbers which should have had a YYNNNN form (e.g. 800010 would be the tenth case filed in 1980) but had been keyed as 8010 with two trailing blanks. This would be converted to 800010.

If a sound justification for conversion could not be made, variables with invalid codes were assigned missing data values.

The next step was to establish a coding system for each variable that would accommodate all years. Four general principles were adopted as part of this process:

- codes that were effective in 1982 would, where appropriate, be used in place of earlier codes. For instance, the code for the middle district of Florida was changed from 30 to 3A in 1972. FCDB records with Code 30 were, therefore, all changed to 3A.
- all docket numbers were converted to a seven byte field, with the format YYNNNN, where YY is the year of filing, and NNNN is the sequence number of the case within the filing office (not the filing district). This was due to accommodate the introduction of this format by the AO in 1983. An exception to this rule is cases without a YY indicator in the first two positions of the docket. These were filed before the YY convention was begun. They were right justified in the seven byte field.
- alphanumeric values used by the AO in some ordinal and categorical variables would be replaced by integer values.
- a two byte field would be created for all variables, partly to accommodate future code expansion, and partly to permit the use of negatives (e.g. -8, -9) as missing data codes.

All variables for all years were then converted to standard codes.

The final step was to rewrite the data for each year into a common format. There is one format for each case type (civil, criminal and appeals). Consequently, programs or analytic procedures that work for any one year within case type will also work for any other year.

IV. The FCDB File Structure and Size

The records for any given fiscal year are grouped by district in a single file for each case type, resulting in three terminations files (Civil, Criminal and Appeals) for each year. Counts of the records in each year of terminations covered by the FCDB are presented in Table IV-1. It is possible that some of these records (perhaps one or two in each year) consist entirely of missing data codes. Researchers should accommodate this possibility in their analysis.

Within each year, the organization of records parallels the structure of the court system (see Table 1 above) — i.e. the data are ordered by circuit, district, office within district and docket number within office.

TABLE 2

*COUNTS OF DISTRICT COURT CASES AND CIRCUIT COURT APPEALS TERMINATED
FY71-FY85 OR PENDING AT START OF FY86*

YEAR	CIVIL	CRIMINAL	APPEALS
FY71 Term.	86,564	50,900	12,427
FY72 Term.	95,182	62,500	13,926
FY73 Term.	98,260	59,026	15,092
FY74 Term.	97,634	56,815	15,364
FY75 Term.	104,784	58,911	16,000
FY76 Term.	110,176	59,512	16,358
FY77 Term.	117,151	57,876	17,784
FY78 Term.	125,914	49,727	17,714
FY79 Term.	143,324	44,567	18,928
FY80 Term.	160,482	39,382	20,887
FY81 Term.	177,975	41,017	25,068
FY82 Term.	189,473	43,325	27,987
FY83 Term.	215,356	46,354	28,662
FY84 Term.	243,113	48,325	31,186
FY85 Term.	269,848	50,421	31,387
FY86 Pend.	254,114	32,620	24,761
TOTALS	2,765,328	847,430	333,531

Civil and Criminal terminations for each district are sorted by office and docket number. The docket number alone is not sufficient to uniquely identify a case because different offices within a district may use the same sequence of docket numbers.

For civil cases, there is a single record within the fiscal year of termination. Multiple parties, whether plaintiffs or defendants, are incorporated into this record, with party specific information being taken from the lead party in each group.

For criminal cases, there is a record for each defendant. Office and docket numbers are the same for these records, so they are distinguished by defendant number and name. This creates identification problems when appeals result from multiple defendant criminal cases because the defendant number is not carried forward to the appeals record. Therefore, defendant name is the only way of ascertaining which of the defendants has appealed.

Circuit Court cases, consisting of appeals from District Court decisions and certain original proceedings, are organized by Circuit and Docket Number. Since each circuit uses a single sequence of docket numbers, no additional identification is needed to uniquely specify a case. All appeals records contain the district, office and docket number of the case being appealed, and it is this that can be used to link an appeal to its district court predecessor.

V. Variables Included In The Federal Court Data Base

The original sources of the information included in the FCDB are the forms that Circuit and District Court Clerks forward to the AO in Washington. Separate forms are used for filing and termination, and, during the life of a case, update information may be transmitted as events occur. The AO subjects the information to certain range and validity checks but makes no substantive changes.

The FCDB contains all information reported to the AO in the following categories:

- Filing Location
- Case Identifiers
- Case Type
- Events and Processing
- Adjudication and Disposition

Certain data items created by the AO after reports are received from clerks' offices have been dropped. These are used by the AO for internal identification and control purposes only, and contain no substantive information about the record to which they apply.

The variables in each of these categories have been carefully screened during the cleaning and editing phases of the FCDB Project, and are now represented either by valid codes or missing data (see the codebooks in Appendices C, D and E for details).

Lists of the variables contained within each of the three general case types — appeals, civil and criminal — are presented in Tables 3A, 3B and 3C respectively.

TABLE 3A

INTEGRATED DATA BASE APPEALS CODEBOOK

-
- | | |
|---------------------------------------|----------------------------------|
| 1. Record Quality Indicator | 28. Transaction Date |
| 2. Appeals Court Circuit | 29. Transaction Code |
| 3. Appeals Court Docket Number | 30. Disposition |
| 4. Reopen Code | 31. By Judicial Action |
| 5. Docket Data (YYMMDD) | 32. Without Judicial Action |
| 6. US as Appellant | 33. Method of Disposition |
| 7. Appellant Name | 34. Opinion/Order |
| 8. US as Appellee | 35. Original Proceeding |
| 9. Appellee Name | 36. Joined Appeal |
| 10. Appeal from Magistrate's Decision | 37. Joined Appeal Docket Number |
| 11. Type of Appeal | 38. Complete Record Filing Date |
| 12. Nature of Original Proceedings | 39. Last Briefs Filing Date |
| 13. In Forma Pauperis | 40. Submission Date |
| 14. Divisional Office | 41. Oral Hearing Date |
| in Appeals Court Circuit | 42. Final Judgement Date |
| 15. Administrative Agency | 43. Case Termination Date |
| 16. Jurisdiction | 44. Misc. to General Docket |
| 17. Nature of Suit | 45. Concur./Dissent. Opinion |
| 18. Offense Code | 46. Probable Cause Decision |
| 19. No Type | for Prisoner Petition |
| 20. District Court Circuit | 47. Who Made Probable Cause |
| decision | |
| 21. District Court District | 48. Single Judge/Full Panel |
| 22. District Court Office | 49. Counsel Appointed |
| 23. District Court Docket Number | 50. Counsel Continued |
| 24. Magistrate Indicator | 51. Counsel Source (if District) |
| 25. Date Filed in District Court | 52. Counsel Source (if Circuit) |
| 26. Date Notice of Appeal Filed | 53. Judge Code #1 |
| 27. Filing Date Used (YYMM) | 54. Judge Code #2 |
-

TABLE 3B

INTEGRATED DATA BASE CIVIL CODEBOOK

1.	Record Quality Indicator
2.	Circuit
3.	District
4.	Filing Office
5.	Filing Docket Number
6.	Filing Date (YYMMDD)
7.	Jurisdiction
8.	Nature of Suit
9.	Origin
10.	Residence
11.	Class Action
12.	Termination Judge
13.	Filing Judge
14.	Trial Date (YYMM)
15.	Demand
16.	Filing Magistrate
17.	County
18.	Style
19.	Termination Date (YYMMDD)
20.	Filing Date Used by AO (YYMM)
21.	Disposition
22.	Termination Magistrate
23.	Procedural Progress
24.	Nature of Judgement
25.	Amount Received
26.	Date Judgement Amount was Received (YYMM)
27.	Judgement for
28.	Magistrate Involvement
29.	Other Involvement
30.	Termination Date Used by AO (YYMM)

TABLE 3C

INTEGRATED DATA BASE CRIMINAL CODEBOOK

1.	Record Quality Indicator
2.	Circuit
3.	District
4.	Filing Office
5.	Filing Docket Number
6.	Defendant Number
7.	Filing Date (YYMM)
8.	Proceeding Code
9.	Filing Offense Code
10.	DuplicateDefendant
11.	Termination Date (YYMM)
12.	Transfer Docket Number
13.	Transfer Circuit
14.	Transfer Defendant Number
15.	Transfer District
16.	Transfer Office
17.	Interval
18.	Offense at Termination
19.	Major Offense Disposition
20.	Counsel
21.	Termination Judge
22.	Observation Code
23.	Sentence Category
24.	Statute
25.	Sentence Type
26.	Prison Term
27.	Probation Term
28.	Fine
29.	Sex
30.	Race
31.	Birth Year
32.	Marital Status
33.	Education
34.	Prior Record
35.	Presentence Investigation
36.	Rule 20 Transfer
37.	Defendant Name
38.	Major Offense Level
39.	Termination Offense Level
