# Technical Addendum to Documentation

# Patent Litigation Data from US District Court Electronic Records (1963-2015)

Richard Miller, Senior Economist, Office of the Chief Economist

Ted Sichelman, University of San Diego School of Law

In March of 2017, the Office of the Chief Economist (OCE) at the US Patent and Trademark Office (USPTO) released two sets of patent litigation data for public use. First, OCE obtained the docket reports on the universe of patent litigation cases in PACER and RECAP and created a dataset for the period 1963-2015. Second, OCE captured the metadata for these cases, which includes information on the case identifier, parties involved, filing date, and district court location. For details regarding the original data release please see Marco, Tesfayesus and Toole (2017), a link to which is also available on the OCE website.

The docket report data that OCE originally provided came in four separate files (i.e. cases, names, attorneys, documents) representing the different sections of the PACER docket reports, while the metadata was provided in the file pacer_cases. One of the main limitations of the initial data release was that it included no information on the patents involved in each lawsuit or on the type of case (for instance, infringement suits brought by patent owners and requests for declaratory judgements regarding patent validity brought by alleged infringers, among others). To remedy this, a research team at the University of San Diego (USD) School of Law reviewed every available initial complaint and related documents in roughly 99 percent of all patent cases filed in U.S. district courts between 2003 and 2016 in order to generate a comprehensive list of litigated patents in those cases. In addition, the team coded the case type for all of these actions, such as infringement, declaratory judgment, false marking, ownership dispute, malpractice, and so forth. This work, which is documented in Schwartz, Sichelman, and Miller (2019) and can also be found on the OCE website, has resulted in the creation of a fifth docket report file (patents) containing the patents-at-issue and case type information (see Figure A-1). The case type variables have also been added to the cases file for the relevant cases.

In addition to this work, the USD research team also updated the existing five files and extended their coverage through the end of calendar year 2016. In this brief addendum to the original documentation, we describe how the underlying data files have changed.

## The cases file

We made a few changes to the cases file as follow:

- The most obvious change is the increase in the number of cases to 81,350. This reflects the addition of the cases from 2016 as well as the addition of cases from the previous years.

- A new *district_id* variable has been added. This is a string variable that takes the following form. The first two characters are the postal abbreviation for the state in which the district court sits. For states with only one district court, the third character is "d" and the fourth character is left blank. For states with multiple district courts the final two characters represent the particular district within the state (e.g., "edca" for the Eastern District of California). Some examples include:
    - "ed" – Eastern District
    - "wd" – Western District
    - "cd" – Central District
    - "md" – Middle District
    - "sd" – Southern District
- The *case_number* variable has been trimmed of information regarding the assigned judge so that it only includes the core docket number (m:yy-xx-nnnnn). We have retained the original PACER case numbers and they are included as the *case_number_raw* variable. As in the original 2017 release, there are a small number of *district_id*/*case_number* pairs that repeat.
- For most of the cases filed since 2003 (54,505 cases), case type information has been added. This describes the type of case (patent infringement, declaratory judgment, false marking, etc.). In a small but significant number of cases, multiple case types (up to three) are identified. Thus we include three case type variables (*case_type_1*, *case_type_2*, and *case_type_3*). Roughly 85 percent of the cases involve patent infringement only, while an additional 8 percent are requests for declaratory judgment. The descriptions of the case_type numeric codes are reproduced from Schwartz, Sichelman, and Miller (2019) in Table A-1.
- The file also includes a new *case_type_note* variable. For cases where the case-type could not be determined with a high enough level of certainty, the *case_type_note* variable takes on a value of "Likely." This occurs in 1,146 of the cases for which a case type is assigned. See Schwartz, Sichelman, and Miller (2019) for more details.

The names file

We made no changes to the names file beyond including information regarding parties to the newly added cases. However, we would like to illustrate an idiosyncrasy of the data. As Marco, Tesfayesus, and Toole (2017) point out in the original documentation, there are situations where more than one name entry occurs per party. As an example, consider the case illustrated in Table A-2.

In this case, one of the defendants is listed as "Wickes Manufacturing Company, a Delaware Corporation, formerly known as, Wickes Manufacturing Company, formerly known as, Gulf and Western Stamping Division of Gulf and Western Manufacturing." Note that each part of the entity's name generates a new observation in the names file. However, the *party_row_count* variable is the same for each entry, indicating that each observation is referring to the same

defendant. We considered cleaning the data in such instances, but decided it was best to document this idiosyncrasy more fully and to allow researchers to decide how they wanted to best deal with it.

The attorneys file

We made only one minor change to the attorneys file beyond including information regarding attorneys representing parties in the newly added cases. In the original release, if an attorney was listed twice or more often on the same case (perhaps representing more than one party) the contact information for the second and later occurrences of the same attorney within the same case would be listed as "(See above for address)." We changed it so that if an attorney is listed multiple times on the same case we use the legitimate contact information listed for the attorney for that specific case to back-fill the contact information for the "See above" observations.

The documents file

As in the cases file, we created a *case_number* variable which trims off the information regarding the assigned judge(s). The original docket number is preserved in the variable *case_number_raw*. The other change to the documents file is the inclusion of a *document_url* variable, which provides a link to the PACER document described in the file.

The pacer_cases file

The *case_number* variable is edited so that it is compatible with the *case_number* variable in the other files. Originally in the pacer_cases file the *case_number* variable was of the form m:**yyyy**-xx-nnnnn. We have edited the variable down to the 13-character version (m:**yy**-xx-nnnnn) found in the other files. The original docket number is preserved in the variable *case_number_raw*. A new variable, *last_pacer_retrieval_date*, has also been added. This lists the date that the PACER data were last gathered by our research team.

The patents file

This is a new file, which provides information regarding the patents involved in cases filed since 2003. The unit of observation is the case-patent pair, which means that there can be multiple observations for a particular case if multiple patents are asserted, challenged, or in some other way involved in that case. Additionally, a particular patent can appear multiple times in the data file if it is involved in multiple cases.

The file contains thirteen variables, most of which can be found in the cases file (See Table A-3). Two of the variables are new. First, the *patent* variable identifies each patent involved in a case. The *patent* variable is a string variable, because design, plant, and re-issued patent numbers are alpha-numeric in nature. The other variable, *patent_doc_type*, identifies the type of patent document involved. In a small number of cases, court disputes involve either patent applications

or even foreign patents rather than granted US patents. The observations in this data file can be matched to other files using the *case_row_id* variable. See Schwartz, Sichelman, and Miller (2019) for a description of how these data were compiled.

## References

Marco, Alan, Asrat Tesfayesus, and Andrew Toole (2017). "Patent Litigation Data from U.S. District Court Electronic Records (1963-2015)." USPTO Economic Working Paper No. 2017-06. Available at SSRN: **https://ssrn.com/abstract=2942295**

Schwartz, David L., Ted Sichelman, and Richard Miller (2019). "USPTO Patent Number and Case Code File Dataset Documentation." USPTO Economic Working Paper No. 2019-05.

**Table A-1: Case Type Code Descriptions**

| Code and Short Description | Long Description |
|---|---|
| 1--Patent Infringement suit non-DJ | Plaintiff is patent holder and sues defendant(s) for infringement of a utility, design, reissue, or plant patent |
| 2—Patent DJ of both Non-Infringement/Invalidity | Accused infringer files declaratory judgment of non-infringement and invalidity/unenforceability |
| 3--Patent DJ of Non-Infringement only | Accused infringer files declaratory of non-infringement only |
| 4--Patent DJ of Invalidity only | Accused infringer files declaratory of invalidity/unenforceability only |
| 5--False Marking | Action for false patent marking of a product |
| 6--Inventorship/Ownership | Action that disputes inventorship or ownership with respect to a patent or patents |
| 7--Pro Se-Incomprehensible | Pro se suit (i.e., filed by a non-attorney) that does not appear to be drafted properly and is difficult to ascertain the patent-related allegation |
| 8--Patent Malpractice/Atty Misconduct | Suit against an attorney for malpractice or misconduct related to patent prosecution, litigation, licensing, etc. |
| 9--Case Opening Error/Other filing error | Case number was opened due to filing error or similar error |
| 10--Non-Patent Case | Complaint has no counts alleging patent causes of action |
| 11--Patent Regulatory/Rule Challenge/Other Administrative Challenge/Patent Term Adjustment | Regulatory/administrative action related to patents, including challenge of USPTO regulation or rule as violating administrative procedure act; claim that USPTO has violated Patent Act procedural rules; request for extension of patent term; challenge to Patent Statute as unconstitutional; or other regulatory/administrative action related to patents |
| 12--Patent Royalty/Licensing Dispute | Dispute among parties to a patent royalty/licensing agreement as to monies owed or related issues |
| 13--Deceptive Invention Promotion | Suit against an "invention promotion company" for engaging in fraudulent or deceptive practices |
| 14--Patent Infringement Foreign Patent | Suit alleging infringement of a patent issued in a foreign country |
| 15--Action collateral to patent case (e.g., independent motion to subpoena documents, quash subpoena, etc.) | Action regarding some collateral issue in a patent case, such as a motion to enforce a subpoena, third-party motion to quash a subpoena; and third-party motion to remove a protective order; and so forth. |

**Table A-2: Example of Party with More than One Row of Name Data**

| case_row_id | party_row_count | party_type | name_row_count | name |
|---|---|---|---|---|
| 46 | 151 | Defendant | 156 | Wickes Manufacturing Company |
| 46 | 151 | Defendant | 157 | a Delaware corporation |
| 46 | 151 | Defendant | 158 | formerly known as |
| 46 | 151 | Defendant | 159 | Wickes Manufacturing Company |
| 46 | 151 | Defendant | 160 | formerly known as |
| 46 | 151 | Defendant | 161 | Gulf and Western Stamping Division of Gulf and Western Manufacturing |

**Table A-3: Variables in the Patents Dataset**

| Variable | Description | Source |
|---|---|---|
| case_row_id | A unique numeric designator assigned by the USPTO to every case in the USPTO Patent Litigation Dataset. | USPTO Patent Litigation Dataset |
| pacer_id | Internal PACER identifier for the case | USPTO Patent Litigation Dataset & PACER |
| case_number | Number assigned by a district court in O:YY-cv-NNNNN format that identifies the office (O) in a particular judicial district, year of filing (YY), and numeric designator (NNNNN) | USPTO Patent Litigation Dataset & PACER |
| district_id | Identifies the district with a standard abbreviation in which the case was filed or the district to which the case was transferred | USPTO Patent Litigation Dataset & PACER |
| nos | Code assigned by the Administrative Office od the Courts (AO) to identify the subject matter of the case. By far, the most common code in the dataset is 830, which is used by the AO to identify patent cases. | USPTO Patent Litigation Dataset & PACER |
| date_filed | Date case was filed | USPTO Patent Litigation Dataset & PACER |
| case_name | Case name as it appears in the USPTO Patent Litigation Dataset. | USPTO Patent Litigation Dataset & PACER |
| case_type_1<br>case_type_2<br>case_type_3 | Identifies specific type of case as described in Table 2 | Manual review of complaints and docket entries available on PACER |
| case_type_note | Identifies cases where there is uncertainty in determining case type. | Manual review of complaints and docket entries available on PACER |
| patent | A patent or patent document identified by number that is at issue in the case. | Manual review of initial complaints, and in some cases, from the amended complaints, counterclaims, docket itself, or other case documents. |
| patent_doc_type | The type of patent document (US patent, application, published application (including PCT filings), or foreign patent. | Manual review (see patent). |

## Figure A-1: Data File Structure (Updated)