Appendix D: Description of the Foreign Priority Data Release

D.1 Data File Included in this Release

This data release consists of one data file called **foreign_priority** that provides all of the information that a user would be able to glean from the "Foreign Priority" tab on PTO's Public PAIR website. The information includes the application number of the subject application, an identifier for the non-US application from which the subject application is claiming priority, the filing date of the non-US application, and the country in which the non-US application was filed. There are 3,788,935 observations on 2,943,998 unique subject applications. An example of the "Foreign Priority" tab is presented in Figure D-1. It shows the "Foreign Priority" tab for application number 10/530,456 which was filed with the PTO in April of 2005. The application, which was never issued as a patent, claims priority from a previous filing with the Japanese Patent Office (JPO) from October of 2002. See Table D-1 for a list of the variables included in the **foreign_priority** data file.

D.2 Variables Included in the Data File

The **foreign_priority** data file includes five variables. The variable *application_number* identifies the subject application which is claiming priority from a foreign application and can be used to link information contained in either of these two data files back to the various other data files that are included in the greater PatEx Research Dataset release.

The variable *foreign_parent_id* identifies the non-US application from which the subject application is claiming priority. The variable *foreign_parent_date* gives the date on which the non-US, parent application was originally filed in the foreign jurisdiction. It is formatted as a numeric variable which is set equal to the difference between the filing date and the first day of January 1960. For instance, if an application was received on 10 January 1960, then the date variable would be equal to 9. For dates prior to 1 January 1960, the date variable takes on negative values. In the Stata version of the data set, the %td display format is embedded, so that the dates display with the following format: ddmmmyyyy. For example, when *foreign_parent_date* is equal to 12,500, it displays in Stata as "23mar1994."

The variables *parent_country_code* and *parent_country* identify the jurisdiction in which the non-US, parent application was filed. The *parent_country_code* variable is coded using the ISO 3166 format.[72]

In Table D-2, we present the most common jurisdictions of original parent filings for non-US, parent applications. Not surprisingly, we find that countries such as Japan, Germany, South Korea, and the United Kingdom are the most common jurisdictions.

Curiously, the United States ranks ninth. Patent Cooperation Treaty (PCT) applications make up roughly 86 percent of the foreign parents coded with the United States as the jurisdiction of original filing. Such applications that do not designate the United States for possible national stage entry are treated as foreign applications, but they should not be coded as being US applications. It is likely that most of these cases amount to either coding errors or errors made by applicants when filing their applications. We have found cases where applicants have claimed foreign priority to regular US applications and where this has been corrected in the published patent, but not in the underlying Public PAIR data. We have also found the following pattern when examining the Public PAIR website. There are many cases where a PCT filing claims the benefit of a previous US application and this appears in the Continuity data for the PCT filing. Then a new regular application is filed as a national stage entry of the PCT filing. The PCT filing appears in the Continuity data for the new regular application while the original US application appears in the foreign priority data for this new regular application. Technically, the original US application should appear as one of the parents of the new application; it should appear in **continuity_parents**, not in **foreign_priority**. We suggest that users proceed with caution when considering applications claiming foreign priority to the United States and make adjustments as necessary.

[72] For more information on ISO 3166 country codes see http://www.iso.org/iso/country_codes.htm.

Table D-1: List of variables included in foreign_priority

Variable Name	Description	Type	Formatting
application_number	Application Number	str14	%-14s
foreign_parent_id	Non-US parent identifier	str17	%-17s
foreign_parent_date	Original filing date of non-US parent	float	%td
parent_country_code	Country of non-US parent code (ISO 3166)	str4	%-2s
parent_country	Country of non-US parent	strL	%-20s

Table D-2: Most common jurisdictions of non-US parent applications

Code	Country/Jurisdiction	Number
JP	Japan	1,809,231
DE	Germany	437,755
KR	South Korea	280,413
GB	United Kingdom	217,684
EP	European Patent Office	171,104
FR	France	149,890
TW	Taiwan	96,928
CN	China	84,092
US	United States	67,545
IT	Italy	54,065
AU	Australia	50,597
SE	Sweden	43,608
CH	Switzerland	34,651
CA	Canada	27,787
NL	Netherlands	23,056

Figure D-1: The Foreign Priority Tab for Application 10/530,456