Appendix C: Description of the Continuity Data Release

C.1 Data Files Included in this Release

This data release consists of two data files that, after matching with the **application_data** file, provide all of the information that an analyst would be able to glean from the "Continuity Data" tab on PTO's Public PAIR website. An example of the "Continuity Data" tab is presented in Figure C-1. Note that the tab is broken out into two panels. The top panel presents information on the parents of the application. Parents are previous applications from which the current application of interest claims benefit. As is illustrated in Figure C-1, a given application can have more than one parent. For example, in this case the regular nonprovisional utility application **10/000,240** claims the benefit of the previous regular nonprovisional application **09/837,917**, which itself claims the benefit of the previous regular nonprovisional application **09/405,294**. The bottom panel presents information on the children of application. In our example we see that three separate applications all claim the benefit of application **10/000,240**. In a sense, however, they are really claiming the benefit of the original application **09/405,294**, as seen in Figure C-2, which shows the continuity data tab for that application. Note also that the applications found in the continuity data tab for **10/000,240** are not the only applications that claim benefit from **09/405,294**. We'll return to this example after we discuss how the data sets are structured.

The two data files provided in this release correspond to the two panels on the "Continuity Data" tab for each application that has either parents or children. The first data file is named **continuity_parents** and it contains information on the application numbers and filing dates of all previous applications from which a subject application claims benefit. If an application has multiple parents, then the file will include multiple observations for that application. This data file includes 6,094,920 observations on 3,838,698 unique subject applications. See Table C-1 for a list of the variables included in **continuity_parents**.

The second data file is **continuity_children** and it contains similar information for all subsequent applications that claim benefit of a given subject application. As was the case with the **continuity_parents** file, the **continuity_children** file includes multiple observations for those applications that have multiple children. This data file includes 5,249,574 observations on 2,612,077 unique subject applications. See Table C-2 for a list of the variables included in **continuity_children**. In the next section we describe the variables in more detail.

C.2 Variables Included in the Two Data Files

Each of the data files contains four variables. Figure C-3 maps the variables to the information that can be found on the Continuity Data tab. The variable *application_number* is common to both the **continuity_parents** and **continuity_children** data sets. This variable identifies the application that has at least one parent (or child) and can be used to link information contained in either of these two data sets back to the various other data sets that are included in the greater PatEx Research Dataset release.

The variable *parent_application_number* is available only in the **continuity_parents** data file and it identifies an application as a parent of the subject application (identified using the *application_number* variable). The variable *child_application_number* is available only in the **continuity_children** data set and it identifies an application as a child of the subject application (again, identified using the *application_number* variable).

The variable *parent_filing_date* is available only in the **continuity_parents** data set. It provides the filing date of the application identified using the *parent_application_number* variable. Likewise, the variable *child_filing_date* is available only in the **continuity_children** data set. It provides the filing date of the application identified using the *child_application_number* variable. Each filing date variable 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 *parent_filing_date* is equal to 12,500, it displays in Stata as "23mar1994."

The final variable, *continuation_type*, is common to the two data sets, but is interpreted differently depending on which data set one is using.[71] In the **continuity_parents** data set, the *continuation_type* variable describes the type of continuation from the parent (identified using *parent_application_number*) to the subject application (identified using *application_number*). In the **continuity_children** data set, the *continuation_type* variable describes the type of continuation from the subject application (identified using *application_number*) to the child (identified using *child_application_number*). The variable can take on any one of 11 values. In Tables C-3 and C-4, we present the counts

[71] It should also be noted that some of the relationships described using the *continuation_type* variables are not necessarily thought of as continuations by PTO. For instance, when a regular nonprovisional application claims the benefit of an earlier provisional application, the latter application is not technically a continuation of the earlier one. This is also usually true when new applications claim the benefit of a prior application filed under the patent cooperation treaty (PCT).

for these values in the **continuity_parents** and **continuity_children** data sets, respectively.

Regardless of the data set examined, claims of benefit from provisional applications tend to be the most common, followed by traditional continuations. Together, they account for roughly 60 percent of the continuations in the **continuity_parents** data set.

C.3 How the Data Files Are Organized

The two data sets are organized to mimic the panels found on the Continuity Data tab on the Public PAIR website. Table C-5 presents the data for application **10/000,240**. The contents of the table can be compared to Figures C-1 and C-3. The **continuity_parents** data set includes two observations for application **10/000,240**, corresponding with the two parents listed in the top panel in Figure C-1, while the **continuity_children** data set includes three observations, corresponding with the thee child applications listed in the bottom panel of the figure.

C.4 Data Irregularities

One oddity in the data is that the vast majority of the national stage entries (denoted by the code "NST") are not represented in the **continuity_children** data set. The main reason for this is that coverage of Patent Cooperation Treaty (PCT) applications is nearly non-existent for the period prior to 1995. Many PCT applications, therefore, are not included in PatEx. When the children of these missing PCT applications (the national stage entries) are included in PatEx, the PCT applications appear in **continuity_parents** as parent applications for these national stage entries. However, because these PCT applications are not in PatEx, we don't see the relationship in the **continuity_children** data set.

There are other inconsistencies in the data that users should keep in mind. For example, beyond the issue with the missing PCT applications, the distributions of the *continuation_type* variable across the two data sets don't match very well. Part of the problem is that there are duplicates of entire lines of data in each data set, although the problem is almost exclusive to the **continuity_children** data set. In that data set there are 78,272 duplicative observations, while in the **continuity_parents** data set there are only 2 duplicative observations. In Table C-6, we compare the frequencies of the different values of the *continuation_type* variable across the two data seta after dropping the duplicates in each. Here the distributions match up better, but there are still some discrepancies. This cannot be explained by inconsistencies in the coding of the *continuation_type* variable across the two data sets. In 99.95 percent of the cases where the parent-child pair can be found in both files (5,133,997 cases), the *continuation_type*

variables match. Most of the differences are driven by the fact that there are significant numbers of parent-child pairs that can be found in one data set but not the other.

For example, there are 956,223 parent-child pairs in **continuity_parents** that are not in **continuity_children**. The vast majority of these (836,626) are national stage entries of PCT applications. However, roughly 97,500 continuations and 21,000 continuations-in-part (CIPs) appear to be missing from **continuity_children** as well. In addition, 32,589 parent-child pairs that are in **continuity_children** are not in **continuity_parents**. Roughly 20,500 of these are cases where the parent is a provisional application, while roughly 10,400 are either continuation applications or CIPs. We suggest that users consider using the **continuity_parents** data set primarily, and supplement it with parent-child pairs that can only be found in the **continuity_children** data set.

Table C-1: List of variables included in continuity_parents

Variable Name	Description	Type	Formatting
application_number	Application Number	str14	%-14s
parent_application_number	Application Number of Parent	str17	%-17s
parent_filing_date	Filing Date of Parent Application	float	%td
continuation_type	Type of Relationship Between Parent and Child Applications	str3	%-3s

Table C-2: List of variables included in continuity_children

Variable Name	Description	Type	Formatting
application_number	Application Number	str14	%-14s
child_application_number	Application Number of Child	str17	%-17s
child_filing_date	Filing Date of Child Application	float	%td
continuation_type	Type of Relationship Between Parent and Child Applications	str3	%-3s

Table C-3: Counts of continuations by continuation type, continuity_parents data file

Value	Description	Frequency	Percent
PRO	Claims the Benefit of a Provisional Application	2,002,590	32.86
CON	Continuation	1,689,776	27.72
NST	National Stage Entry	938,002	15.39
CIP	Continuation in Part	794,673	13.04
DIV	Divisional Continuation	617,963	10.14
REI	Re-Issue	23,836	0.39
REX	Re-Examination	15,402	0.25
?	No Data	12,494	0.2
SUB	Substitute Application	130	0
SER	Supplemental Examination	53	0
RIC	Unknown	1	0
	Total	6,094,920	100.00

Table C-4: Counts of continuations by continuation type, continuity_children data file

Value	Description	Frequency	Percent
PRO	Claims the Benefit of a Provisional Application	2,064,484	39.33
CON	Continuation	1,601,690	30.51
CIP	Continuation in Part	808,794	15.41
DIV	Divisional Continuation	620,909	11.83
NST	National Stage Entry	101,390	1.93
REI	Re-Issue	24,202	0.46
REX	Re-Examination	15,434	0.29
?	No Data	12,509	0.24
SUB	Substitute Application	108	0
SER	Supplemental Examination	53	0
RIC	Unknown	1	0
	Total	5,249,574	100.00

Table C-5: Continuity data for Application 10/000,240

Panel: continuity_parents

application_number	parent_application_number	parent_filing_date	continuation_type
10000240	9405294	23sep1999	CON
10000240	9837917	19apr2001	CON

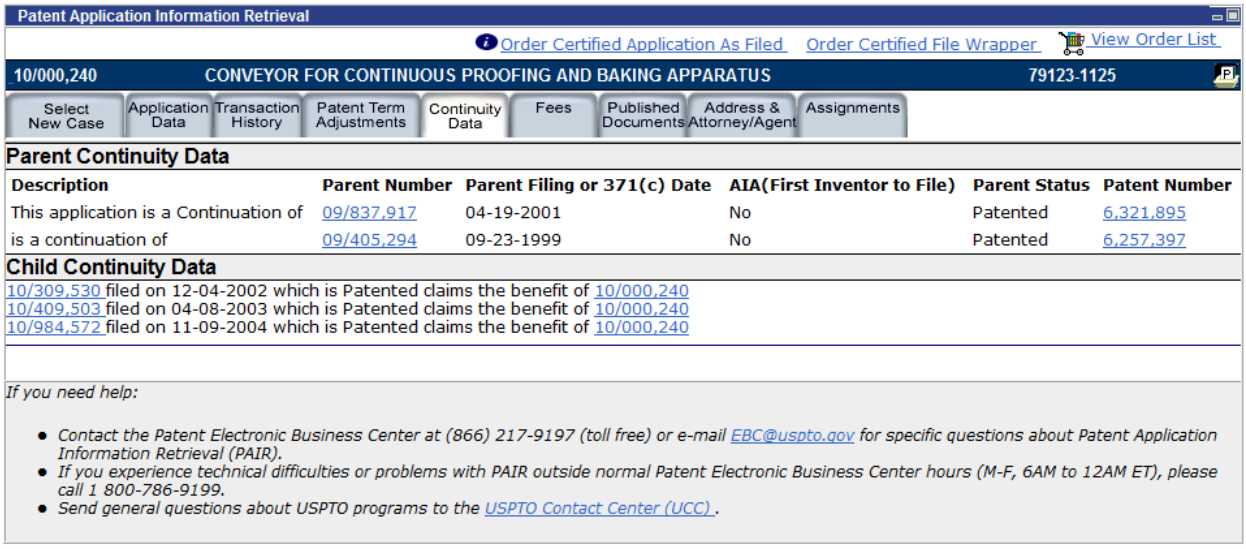
Panel: continuity_children

application_number	child_application_number	child_filing_date	continuation_type
10000240	10409503	08apr2003	CON
10000240	10309530	04dec2002	CON
10000240	10984572	09nov2004	CON

Table C-6: Comparing the distribution of the *continuation_type* variable across data files after removing duplicates

		Frequency in CONTINUITY_...	
Value	Description	PARENTS	CHILDREN
PRO	Claims the Benefit of a Provisional Application	2,002,590	2,022,356
CON	Continuation	1,689,775	1,595,906
NST	National Stage Entry	938,002	101,305
CIP	Continuation in Part	794,672	780,477
DIV	Divisional Continuation	617,963	619,236
REI	Re-Issue	23,836	23,962
REX	Re-Examination	15,402	15,405
?	No Data	12,494	12,493
SUB	Substitute Application	130	108
SER	Supplemental Examination	53	53
RIC	Unknown	1	1
	Total	6,094,918	5,171,302

Figure C-1: The Continuity Data Tab for Application 10/000,240



Patent Application Information Retrieval

Order Certified Application As Filed | Order Certified File Wrapper | View Order List

10/000,240 CONVEYOR FOR CONTINUOUS PROOFING AND BAKING APPARATUS 79123-1125

Select New Case | Application Data | Transaction History | Patent Term Adjustments | Continuity Data | Fees | Published Documents | Address & Attorney/Agent | Assignments

Parent Continuity Data

Description	Parent Number	Parent Filing or 371(c) Date	AIA(First Inventor to File)	Parent Status	Patent Number
This application is a Continuation of	09/837,917	04-19-2001	No	Patented	6,321,895
is a continuation of	09/405,294	09-23-1999	No	Patented	6,257,397

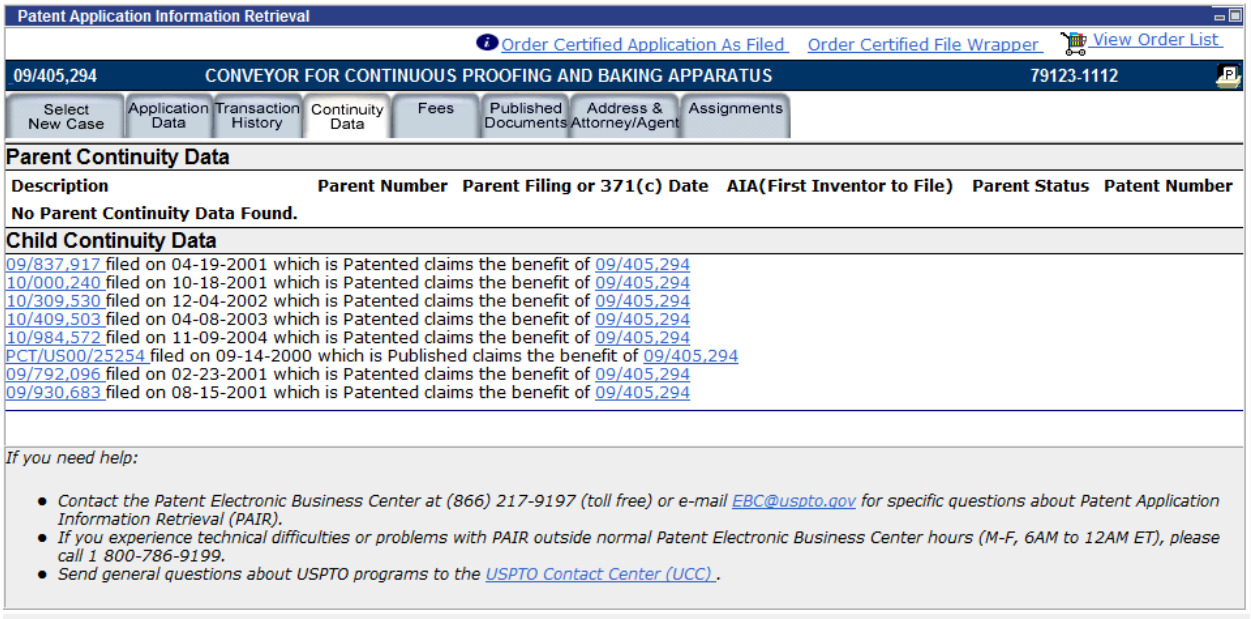
Child Continuity Data

10/309,530 filed on 12-04-2002 which is Patented claims the benefit of 10/000,240
10/409,503 filed on 04-08-2003 which is Patented claims the benefit of 10/000,240
10/984,572 filed on 11-09-2004 which is Patented claims the benefit of 10/000,240

If you need help:

- *Contact the Patent Electronic Business Center at (866) 217-9197 (toll free) or e-mail EBC@uspto.gov for specific questions about Patent Application Information Retrieval (PAIR).*
- *If you experience technical difficulties or problems with PAIR outside normal Patent Electronic Business Center hours (M-F, 6AM to 12AM ET), please call 1 800-786-9199.*
- *Send general questions about USPTO programs to the USPTO Contact Center (UCC) .*

Figure C-2: The Continuity Data Tab for Application 09/405,294



Patent Application Information Retrieval

Order Certified Application As Filed | Order Certified File Wrapper | View Order List

09/405,294 CONVEYOR FOR CONTINUOUS PROOFING AND BAKING APPARATUS 79123-1112

Select New Case | Application Data | Transaction History | Continuity Data | Fees | Published Documents | Address & Attorney/Agent | Assignments

Parent Continuity Data

Description	Parent Number	Parent Filing or 371(c) Date	AIA(First Inventor to File)	Parent Status	Patent Number
No Parent Continuity Data Found.					

Child Continuity Data

09/837,917 filed on 04-19-2001 which is Patented claims the benefit of 09/405,294
10/000,240 filed on 10-18-2001 which is Patented claims the benefit of 09/405,294
10/309,530 filed on 12-04-2002 which is Patented claims the benefit of 09/405,294
10/409,503 filed on 04-08-2003 which is Patented claims the benefit of 09/405,294
10/984,572 filed on 11-09-2004 which is Patented claims the benefit of 09/405,294
PCT/US00/25254 filed on 09-14-2000 which is Published claims the benefit of 09/405,294
09/792,096 filed on 02-23-2001 which is Patented claims the benefit of 09/405,294
09/930,683 filed on 08-15-2001 which is Patented claims the benefit of 09/405,294

If you need help:

- *Contact the Patent Electronic Business Center at (866) 217-9197 (toll free) or e-mail EBC@uspto.gov for specific questions about Patent Application Information Retrieval (PAIR).*
- *If you experience technical difficulties or problems with PAIR outside normal Patent Electronic Business Center hours (M-F, 6AM to 12AM ET), please call 1 800-786-9199.*
- *Send general questions about USPTO programs to the USPTO Contact Center (UCC) .*

Figure C-3: Mapping the variables to the Continuity Data Tab