# Patent application bibliographic data documentation

Bhaven N. Sampat
Assistant Professor
Mailman School of Public Health
Columbia University
bns3@columbia.edu

August 3, 2011

## 1 Summary

This file includes information on inventors for U.S. patent applications published between March 15, 2001 and December 31, 2010, including their names and locations. The data were extracted from the USPTO's *Cassis* BIB Applications DVD.

The dataset includes the application number, inventor name, inventor sequence, and city, state, country information. Each observation is an application-inventor pair, i.e. if there are N inventors on an application that application is repeated N times in the data.

# 2 Codebook

```
. codebook
```

---

appnum (unlabeled)

---

```
                type:  string (str8)
        unique values:  2624383                  missing "":  0/6517914
            examples:  "10615768"
                       "11229126"
                       "11819954"
                       "12402170"
```

---

inv_seq Inventor Sequence

---

```
                type:  string (str1)
        unique values:  6                        missing "":  0/6517914
           tabulation:  Freq.  Value
                       2.6e+06 "1"
                       1.7e+06 "2"
                       1.1e+06 "3"
                       6.1e+05 "4"
                       3.3e+05 "5"
                       1.8e+05 "6"
```

---

lname Inventor Last Name

---

```
                type:  string (str15)
        unique values:  534773                   missing "":  0/6517914
            examples:  "Elwart"
                       "Kang"
                       "Mudda"
                       "Siedlarz"
             warning:  variable has embedded blanks
```

---

fname Inventor First Name

---

```
                type:  string (str13)
        unique values:  329591                   missing "":  18191/6517914
            examples:  "Edward"
                       "John"
                       "Michael"
                       "Siegfried"
             warning:  variable has embedded blanks
```

---

city Inventor City

---

```
                type:  string (str13)
        unique values:  176008                   missing "":  30618/6517914
            examples:  "Duluth"
                       "Kawasaki"
```

```
                        "Osaka"
                        "South Bend"
              warning:  variable has embedded blanks
-------------------------------------------------------------------------------
state                                                          Inventor State
-------------------------------------------------------------------------------

                 type:  string (str2)
         unique values:  57                      missing "":  3476831/6517914
              examples:  ""
                        ""
                        "CA"
                        "MO"
-------------------------------------------------------------------------------
country                                                      Inventor Country
-------------------------------------------------------------------------------

                 type:  string (str2)
         unique values:  184                     missing "":  2818187/6517914
              examples:  ""
                        ""
                        "DE"
                        "JP"
```

# 3 Data construction notes

Application numbers can be linked to bibliographic information in the `applications20012010_bib` file.

Note that state and country information are individually missing for many applications. This is because country is set to missing if the country is "US" and state is missing if the inventor is from a country other than the US. Collectively, only a small share of the observations (15,409) have both country and state missing.