Document Title:	Federal Justice Statistics Program Data Linking System
Author:	Jessica A. Kelly
Document No.:	239536
Date Received:	September 2012
Award Number:	2010-BJ-CX-K079

Opinions or points of view expressed are those of the author(s) and do not necessarily reflect the official position or policies of the U.S. Department of Justice.

TECHNICAL REPORT AUGUST 2012

Federal Justice Statistics Program Data Linking System

Jessica A. Kelly, Urban Institute

URBAN INSTITUTE
Justice Policy Center

URBAN INSTITUTE
Justice Policy Center

2100 M Street NW
Washington, DC 20037
www.urban.org

The reviews expressed are those of the authors and should not be attributed to the Urban Institute, its trustees, or its funders.

This project was supported by Grant No. 2010-BJ-CX-K079 awarded by the Bureau of Justice Statistics, Office of Justice Programs, U.S. Department of Justice. Points of view in this document are those of the author and do not necessarily represent the official position or policies of the US Department of Justice.

Acknowledgements

The author extends gratitude to Mark Coggeshall formerly of the Urban Institute for his initial conception of the dyad links, as well as William Adams, Julie Samuels of the Urban Institute and Avinash Bhati of Maxarth LLC, for their guidance. Additionally, the author would like to thank Samuel Taxy and Jennifer Yahner of the Urban Institute for their thoughtful comments in their review of the final document.

About the Author

Jessica A. Kelly is a Senior Programmer/Analyst and the chief developer of the Dyad Linking System. Ms. Kelly has been a member of the Federal Justice Statistics Program (FJSP) team since 2007. In addition, she has experience in database development and programming in several computing languages including C++, C# and SAS. Ms. Kelly holds a BS in mathematics magna cum laude from Towson University.

Table of Contents

List of Tables and Figures

I. Introduction

This technical report provides a detailed description of the Federal Justice Statistics Program (FJSP) dyad linking system, focusing on the system's methodology. It is designed for users of the FJSP dyad linking system as well as a more general audience interested in data linking methodologies. The report is divided into several sections. The first sections provide general background information about the Federal Justice Statistics Program as a context for readers unfamiliar with the FJSP along with a brief description of an earlier linking methodology. Next, the new linking methodology – the dyad linking system – is described in detail. Results of the dyad linking are then presented and compared to the results from the previous system. The appendices provide more in-depth information about the implementation of the Jaro-Winkler algorithm, detailed results, and comparisons to the first generation linking system.

II. Background

The Federal Justice Statistics Program, funded by the Bureau of Justice Statistics (BJS) and operated under a cooperative agreement by the Urban Institute (UI), provides comprehensive information about suspects and defendants processed in the federal criminal justice system. The goal of the FJSP is to provide uniform case processing statistics across all stages of the federal criminal justice system, including arrest, prosecution, pretrial release, adjudication, sentencing, incarceration, and supervision.[1] The FJSP collects administrative data from six federal criminal justice agencies: the Drug Enforcement Administration (DEA), the United States Marshals Service (USMS), the Executive Office for United States Attorneys (EOUSA), the Administrative Office of the U.S. Courts (AOUSC), the United States Sentencing Commission (USSC), and the Bureau of Prisons (BOP). Data files are received from these agencies in a variety of formats each year and are converted into a comprehensive, standardized database that forms the backbone of the FJSP.

The core concept of the FJSP data system is the Standard Analysis File (SAF). Raw data received from each agency (typically pertaining to a specific stage of case processing) is first converted into a standardized format (SAF) in terms of offense classification, reporting period, and unit of analysis. The person-case is the typical unit of analysis of the SAFs. For example, an individual involved in multiple cases will be counted in each case in the AOUSC SAFs; similarly, codefendants in a single case will each be considered distinct units in the AOUSC SAFs. This is true for most of the SAFs; however, there are two exceptions where such distinctions are not made: the USSC SAFs record sentencing events on a particular date as its unit of analysis and the Bureau of Prisons SAFs report only on the movements of prisoners.

Up to three distinct fiscal year SAFs are created for each stage—called the "IN", the "OUT", and the "STK". The IN SAF refers to an entering cohort during a particular fiscal year. The OUT SAF refers to the exiting cohort during that fiscal year. And the STK cohort refers to the stock (outstanding) at the end of the fiscal year.

It is important to note that while the linking process takes advantage of person-level variables such as docket number, defendant name and date fields, these identifying variables are redacted or sanitized in the SAF files to NACJD. Thus, since users of the publicly available data at NACJD do not have access to these variables, they cannot use them to perform links themselves in the same manner; however, through the FJSP dyad link files, which provide cross-walks of offender records across two data sources via sequential ID numbers that are appended to each agency dataset in the standard analysis files, users have the ability to link records across stages/agencies. .

[1] Additional information about the FJSP, including annual statistical tables and an online statistics tool, are available on the BJS website: http://bjs.ojp.usdoj.gov/fjsrc/. For more information about obtaining FJSP SAFs and linking files, which are available on a restricted use basis, please check the NACJD website: http://www.icpsr.umich.edu/icpsrweb/content/NACJD/guides/fjsp.html.

Despite the comprehensiveness of the coverage of the FJSP, there is one limitation in the structure of the data when assembled as a series of SAFs. Each SAF is a stand-alone dataset pertaining to a particular year that cannot be readily linked to another. Suspects, defendants, and offenders progress through the criminal justice system and sometimes suspects investigated in a particular year may be tried and adjudicated the next year and enter a BOP facility the following year. Hence, stand-alone SAFs on their own do not allow a user the ability to track person-cases through the various stages of the federal criminal justice system. In the late 1990s, the Urban Institute and BJS recognized the value of a system that would enable users to track persons through the system. Over the course of several years, UI staff developed the first generation of such a system, described in the next section, and then greatly improved this system through the use of paired-agency (dyad) links, which are described subsequently and comprise the primary focus of this technical report.

III. First Generation Link System

The original Link Index System (often referred to as either "LIS" or "LIF", for "Link Index File") was designed for comprehensive coverage of all stages of the federal criminal justice system and for scalability—i.e., if and when new agencies decided to contribute data to the FJSP, they could easily be included in the linking system. Unfortunately, preserving these desirable features – comprehensiveness and scalability – did not permit UI staff to take advantage of the full set of personal identifiers shared by a given pair of agencies representing adjacent stages that, if used, could have optimized link rates.

The linking variables included in the first generation system were the Federal Judicial Circuit and District, the Court Docket Number, and the Suspect/Defendant Name. These variables were selected primarily because they were readily available in nearly all of the SAFs included in the LIS. As a result, this meant the key case processing variables like dates (e.g., adjudication date) and other identifiers (e.g., Social Security Number or FBI Number) could not be included in the linking algorithm, because they were not consistently available across all agency datasets. In addition, the first generation system was not designed to be directional—a combination of circuit/district, court docket number, and defendant name was used to create a key for each observation in each cohort and all cohort keys across all years were put into an algorithm that made pair-wise matches in an all-to-all fashion. Key conditioning information (case processing exit points) was not considered when computing match rates. For example, defendants who had their cases dismissed (or who were found not guilty) would not be tracked in the USSC data (sentencing stage) even though they were to be found in the AOUSC (adjudication stage). Similarly, guilty offenders sentenced to only probation (from the USSC data) would not be found in the BOP IN cohort (entering BOP facilities). Ignoring these conditioning rules made it difficult to assess the quality of the first generation LIS and hard for users to distinguish false negatives (links that were not found but should have) from true negatives (links that should not have been found). Furthermore, linkage rates across pairs of agencies were simply not as high as they could have been due to the limited set of linking criteria variables that were required in an all-to-all based linking system. It was recognized that improved link rates could be achieved between adjacent pairs of agencies by using additional linking variables (e.g., processing event dates, and other identifiers, such as FBI Number) that the two agencies shared, but which were not necessarily common across all agencies. Finally, the dissemination vehicle for the first generation system was cumbersome: all links across all stages and all years (1994-present) were stored in a single, large link index file. Users interested in assessing links across just two stages—e.g., AOUSC and USSC—still needed to access, read, and process the full file. As more data years and stages were added to the system, the file became increasingly larger and more elaborate.

IV. Dyad Linking System

Given the limitations noted above, as well as BJS's desire for a more accurate and user-friendly linking system that was based on recent advances in algorithmic matching, the Urban Institute developed a paired-agency, i.e. dyad, linking system. The modified system has several important new features that are outlined below. BJS currently makes this linking system available to users, on a restricted-use/approval basis, through the Inter-university Consortium of Political and Social Research (ICPSR) at the University of Michigan.

A. Dyad-based System

The new FJSP linking system was designed as a dyad-based system. That is, links are established between pairs of agency files (or "dyads") from adjacent stages of case processing. There are several advantages of developing a dyad-based system. Primarily, variables that can be used to establish links could now be selected one dyad-pair at a time, and did not need to be common across all agencies. Inter-agency links provide linkages between two agencies (see Table IV-1, below), and intra-agency links provide linkages within the same agency across cohorts (see Table IV-2, below). This dyad-based approach greatly increased the ability to select stage- and agency-specific identifiers or demographic variables that may not exist in every possible dyad.

Table IV-1 Inter-Agency Links

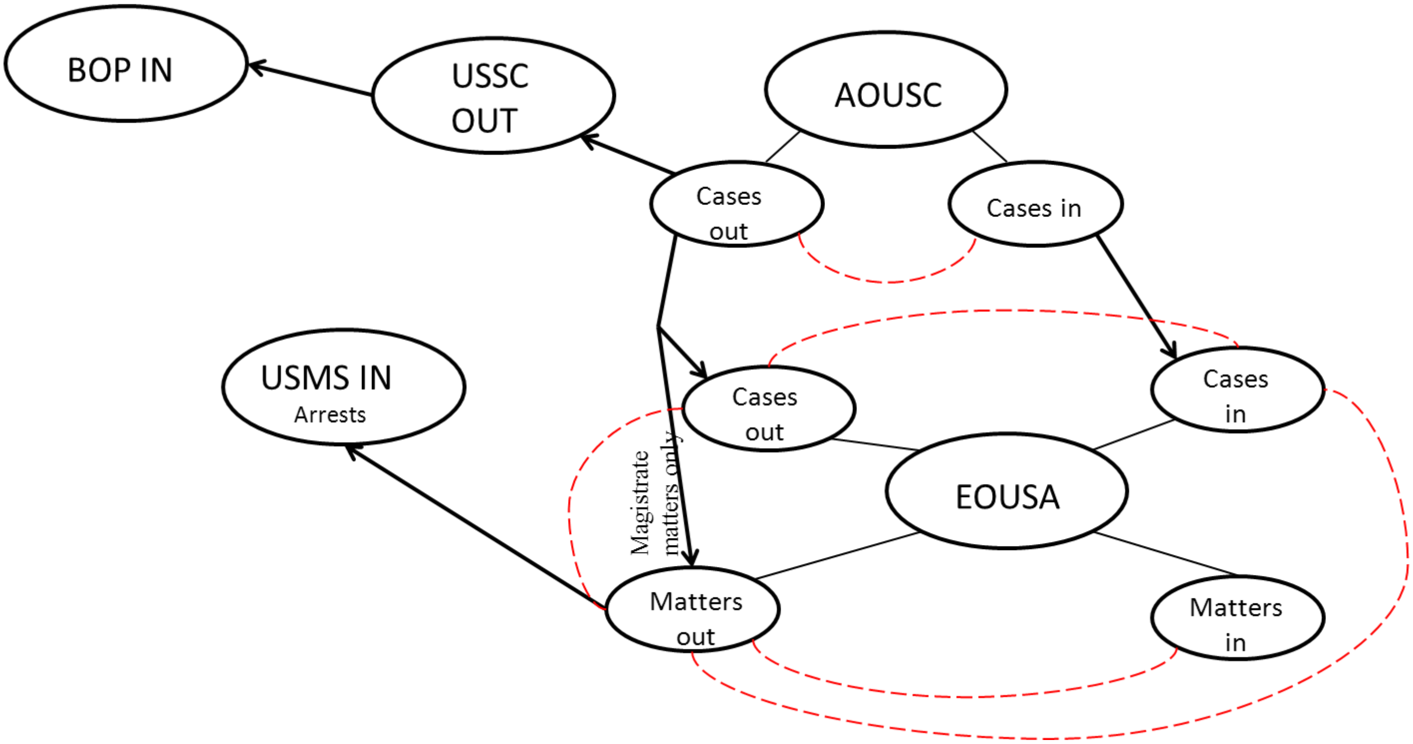
EOUSA MATTERS OUT (Suspects in criminal matters concluded by U.S. attorneys)	USMS IN (Persons arrested for suspected violations of federal law and booked by the U.S. Marshals Service)
AOUSC CASES IN (Defendants in criminal cases filed in U.S. district court)	EOUSA CASES IN (Defendants in criminal cases filed in U.S. district court)
AOUSC CASES OUT (Defendants in criminal cases concluded in U.S. district court)	EOUSA CASES OUT & EOUSA MATTERS OUT (Magistrate records only) (Defendants in criminal cases filed in U.S. district court and Suspects in matters disposed by U.S. Magistrates)
AOUSC CASES OUT (Defendants in criminal cases concluded in U.S. district court)	USSC OUT (Offenders sentenced pursuant to the Federal Sentencing Reform Act of 1984)
USSC OUT (Offenders sentenced pursuant to the Federal Sentencing Reform Act of 1984)	BOP IN (Prisoners entering federal prison)

Table IV-2 Intra-Agency Links

AOUSC CASES IN (Defendants in criminal cases filed in U.S. district court)	AOUSC CASES OUT (Defendants in criminal cases concluded in U.S. district court)
EOUSA CASES IN (Defendants in criminal cases filed in U.S. district court)	EOUSA CASES OUT (Defendants in criminal cases filed in U.S. district court)
EOUSA MATTERS OUT (Suspects in matters disposed by U.S. Magistrates)	EOUSA CASES OUT (Defendants in criminal cases filed in U.S. district court)
EOUSA MATTERS OUT (Suspects in criminal matters concluded by U.S. attorneys)	EOUSA CASES IN (Defendants in criminal cases filed in U.S. district court)
EOUSA MATTERS IN (Suspects in criminal matters opened by U.S. attorneys)	EOUSA MATTERS OUT (Suspects in criminal matters concluded by U.S. attorneys)

The figure below (Figure IV-1) shows all links that have been completed between stages/agencies with solid line arrows. The dotted lines show link pairs within the same agency, between cohorts (intra-agency links).

Figure IV-1 Diagram of Inter- and Intra-Agency Links



B. *Conditioned Links*

Another new feature of the linking system is that links are based on a conditional subset of records. This is possible because of the dyad-based approach. For example, when assessing the link between the AOUSC and the USSC data, records pertaining to dismissed cases are dropped from the AOUSC data prior to matching, leading to a more reasonable starting point for the matching exercise. Table IV-3, see below, shows the screening conditions that were applied to each dyad. Note that screening conditions apply only within a specific dyad. The AOUSC data, for example, have screening conditions applied to them when linked to USSC data, but not when linked to EOUSA data.

Table IV-3 Screening Rules Used by Dyad

Dyad	Screening Rules	
EOUSA MATTERS OUT/USMS IN	USMS:	Material witnesses and supervision violations (tigron = 111 and 112) are not included.
	EOUSA:	None
AOUSC IN/EO IN	AOUSC:	None
	EOUSA:	None
AOUSC OUT/EOUSA OUT	AOUSC:	None
	EOUSA:	None
AOUSC OUT/USSC OUT	AOUSC :	Only defendants convicted (outcome = 1, 2, 3 or 4) are retained.
	USSC:	None
BOP IN/USSC OUT	BOP:	Only defendants sentenced to prison for new U.S. district court commitments (howcomt = 101) are retained.
	USSC:	Only those sentenced to prison (For years prior to 1998, TotDays > 0 or TotPrisn > 0. For 1998 and forward, SentImp = 1, 2) are retained.
AOUSC INTRA Links	AOUSC:	None
EOUSA INTRA Links	EOUSA:	None

NOTES: Other differences between the data sets may still be present but are not systematically screened out. For example, the AOUSC data contain some juvenile defendants that will not be present in the USSC data. These are discussed further in the Results section

C. Blocking and Matching Variables

The new methodology uses two sets of variables: those used for blocking and those used for matching. Blocking variables are used to create bins within which matching is performed using the Jaro-Winkler algorithm. This matching uses name as the main matching variable, as described below in the “Jaro-Winkler Matching Algorithm” section.

A block is calculated by concatenating a set of blocking variables together into a string. As we are linking two different datasets, it is integral that the blocking variables on each dataset have identical coding schemes to ensure that we are comparing like values. When two datasets A and B are linked, one dataset, B, is read into memory in its entirety. As observations in B are read, the block is calculated. For example, if district and docket number(*) are used as the blocking variables then the block might have the value “70200200043”, where the district is equal to “70” and docket number is equal to “200200043”.

After all of the observations in B are read, the records are sorted by block. Then, each observation in A is read one-by-one. As they are read, the block is calculated and the record in A is compared to all observations in B within the same block. The best links between A and B are kept, and if the Jaro-Winkler score meets or exceeds a given threshold, then the observations are considered a match.

The algorithm iterates over several different blocks in order to maximize the number of links found. A more detailed discussion of each dyad can be found in the “Processing Details” section. Table IV-4, below, shows a final list of variables used for blocking and their definitions, and is followed by Table IV-5, which shows in which dyads the variables are used.

*Docket numbers are standardized prior to linking

Table IV-4 Definition of All Blocking Variables Used By Agency

Agency	Variable	Description
USMS	ARDATE/ARRST_DT	Arrest date
	CRTCNUM	Court case number
	DIST	District code
AOUSC	DCKET_YR	Docket year
	DCKT_NUM	Docket number
	DEFEND	Defendant number
	DISP_MM	Disposition month
	DISP_YY	Disposition year
	DISTRICT	District code
	FIL_MM	Case filing month
	FIL_YY	Case filing year
	FILEMAG	U.S. magistrate flag
	SENT_MM	Sentencing month
	SENT_YY	Sentencing year
	TRM_YM	Termination year and month
	TRMJUDG1-4	Judge identifiers
EOUSA	ARREST_YM	Arrest year month
	CLAIM	Claim number
	COURTNBR	Court case number
	DISP_YM	Disposition year and month
	DISTRICT	District code
	FIL-YM	Case filing year and month
	LIONS	Legal Information Office Network System (LIONS) Number
	MAGFLAG	Matter concluded by magistrate flag
	RCV_YM	Year and month matter received
	TERM_YM	Case/Matter termination year and month
USSC	DEFSSN	Defendant Social Security Number
	DISTRICT	District code
	DOCKETID	Docket number
	FBINUM	Defendant FBI Number
	JUDGE	Judge identifiers
	MARSLNUM	Defendant Marshals Number
	SENTDATE	Sentencing date
BOP	DOCKTNO	Docket number
	FBINUM	Inmate FBI Number
	REGNO	Inmate Register Number
	SSNNUM	Inmate Social Security Number

Table IV-5 Blocking Variables Used by Dyad

EOUSA Matters OUT/USMS IN	USMS: ARDATE, CRTCNUM, DIST, first three letters of first and last names EOUSA: ARREST_YM, RCV_YM, COURTNBR, DISTRICT, first three letters of first and last names
AOUSC IN/EOUSA IN; AOUSC OUT/ EOUSA OUT	AOUSC: DCKT_YR, DCKT_NUM, DISTRICT, FILEMAG, DISP_YY, DISP_MM, FIL_YY, FIL_MM, TRM_YM, first three letters of first and last names EOUSA: COURTNBR, MAGFLAG, DISP_YM, FIL_YM, TERM_YM, DISTRICT, first three letters of first and last names
AOUSC OUT/USSC OUT	AOUSC: DISTRICT, TRMJUDG1-4, DCKET_YR, DCKT_NUM, SENT_YY, SENT_MM, first three letters of last name USSC: DISTRICT, JUDGE, DOCKETID, SENTDATE, first three letters of last name
USSC OUT/BOP IN*	USSC: MARSLNUM , DEFSSN, FBINUM, DOCKETID BOP: REGNO, SSNNUM, FBINUM, DOCKTNO
AOUSC INTRA LINKS	DISTRICT, DCKET_YR, DCKT_NUM, DEFEND
EOUSA INTRA LINKS	DISTRICT, LIONS, CLAIM, DISP_YM, TERM_YM, FIL_YM, DISTRICT, first three letters of last name

**NOTE: MARSLNUM in USSC is identical to BOP REGNO.*

With the exception of the BOP/USSC link, name is the only matching variable used. The BOP/USSC dyad makes use of other personal identifiers, Social Security Number, Marshals Number and FBI Number. Name is disregarded when using Marshals Number, FBI Number or Social Security Number and it is assumed that if these values match, then the records in BOP and USSC refer to the same person.[2] In other dyads, personal identifiers may be available, but only in one dataset (e.g. USSC has Social Security Number, but AOUSC does not), so they are not used in the dyad.

D. Jaro-Winkler Matching Algorithm

When comparing two observations in a block, the new paired-agency linking system uses the Jaro-Winkler distance as a measure of similarity between two names.[3] The Jaro-Winkler distance is normalized such that a result of one indicates an exact match and a result of zero indicates no similarity. The Jaro-Winkler distance is a modification of the Jaro distance. The Jaro measure is the weighted sum of percentage of matched characters from each string and transposed characters. Winkler increased this measure for matching initial characters, and then rescaled it by a piecewise function, whose intervals and weights depend on the type of string (name, address, etc.). In short, the Jaro-Winkler distance algorithm yields a quick but flexible matching approach for strings that are mostly the same but may vary in arbitrary ways.

Though a few of the source datasets contain additional person-level identifiers such as Social Security Number, most only have name. As such, the Jaro-Winkler distance is the best choice (versus other linkage techniques) in determining if two records match.[4]

Additional details of how this algorithm was implemented can be found in the Appendix.

[2] See the C++ Processing section for the BOP/USSC Dyad for more details.

[3] The first generation linking system, in most cases, tried to link based on an exact match for name over several iterations: using full name, removing suffixes, dropping middle initial, and using only a substring of the full name.

[4] Peter Christen. 2006. "A Comparison of Personal Name Matching: Techniques and Practical Issues." Available: http://cs.anu.edu.au/techreports/2006/TR-CS-06-02.pdf

E. Thresholds

As described above, the Jaro-Winkler algorithm depends on a threshold to determine if two names are similar enough to be considered a match. In general, these thresholds are set such that the fewer observations in a block, the lower the threshold value. Similarly, the more observations contained in a block, the higher the threshold. When setting the value of the threshold, we examined the results and if it appeared that too many incorrect links were being set, the value was raised. Likewise, if it seemed that too many links were missed, the value was lowered. Further refinements to these values could improve potentially results.

V. Processing Details

The linking process consists of two stages. The first portion of the process is the creation of the files used in each dyad link. The second stage uses these files, calculates the links for a particular dyad, and writes output files.

A. Stage 1: SAS Data Prep

This stage creates a single file for each agency source and cohort to be used for fiscal years 1994 – 2009. Not only does this step append all years together, but, more importantly, it also standardizes variables both within the cohort and across agencies. For example, a variable in the AOUSC data may change from character to numeric, or vary in length, across years. Alternatively, a variable such as district may be coded differently in the AOUSC and USSC SAFs. These must be recoded to an identical scheme prior to attempting to link based on this information.

Most importantly, name fields must be cleaned, standardized and parsed into first, middle and last name. While some source datasets have separate fields for last, first and middle names, others do not, as illustrated below in Table V-1.

Table V-1 Name Variables Available by Agency

Agency	Variable	Description
AOUSC	NAME	First, middle, and last name (includes corporation names)
BOP	INAME_F	Prisoner first name
BOP	INAME_L	Prisoner last name
BOP	INAME_M	Prisoner middle name
EOUSA Cases, Matters	FIRST_NAME	First and middle name
EOUSA Cases, Matters	LAST_NAME	Last name (includes corporation names)
EOUSA Cases, Matters	NAME*	Last name (includes corporation names)
USMS	NMFNAME	Prisoner first name
USMS	NMLNAME	Prisoner last name
USMS	NMMNAME	Prisoner middle name
USSC	DNAME_F	First name
USSC	DNAME_L	Last name
USSC	DNAME_M	Middle name
USSC	DNAME_S	Name suffix (Jr, Sr, etc.)

** EOUSA NAME field used in years prior to 2004. From 2004 forward, separate first and last name fields are used.*

Additionally, even if both agencies in a pair do have separate name fields, we cannot assume that they will record even last names in a common manner. For example, a person might have the full name “JOHN JACOB SMITH-JONES”. The data can be stored in any number of ways:

Table V-2 Possible Name Storage Variations

Last Name	First Name	Middle Name
SMITH-JONES	JOHN	JACOB
SMITH JONES	JOHN	JACOB
SMITHJONES	JOHN	J
JONES	JOHN	SMITH

Further complicating matters is the question of how to divide a single name field up into its component parts. For names containing a comma, this problem is more straightforward; the word(s) preceding the comma are stored as the last name, the word following the comma is stored as the first name and the next word (if it exists and is longer than three characters) is stored as the middle name. In cases where there is no comma, the first word is stored as last name, the next as first name and the third (if it exists and is longer than three characters) is stored as the middle name. Finally, all components are then compressed (removing spaces) and stripped of any special characters. This removes cases such as a last name being “VAN WINKLE” in one dataset and “VANWINKLE” in another.

There is one minor caveat when a name has four name parts (“SMITH, JOHN JACOB JONES”, for example). This occurs often in the AOUSC data; in the EOUSA data, however, the name would likely be stored as “JONES-SMITH, JOHN JACOB”. As a result, if a name has an additional word following what is used as the middle name, it will be stored as a prefix to the last name:

Table V-3 Examples of Name Parsing

In SAF	Calculated Fields		
Full Name	Last Name	First Name	Middle Name
SMITH, JOHN	SMITH	JOHN	
SMITH, JOHN JACOB	SMITH	JOHN	JACOB
SMITH-JONES, JOHN JACOB	SMITHJONES	JOHN	JACOB
SMITH, JOHN J*	SMITH	JOHN	
SMITH, JOHN JACOB JONES	JONESSMITH	JOHN	JACOB
JOHN JACOB SMITH	JOHN	JACOB	SMITH

**NOTE: As discussed above, middle initial is disregarded by the linking algorithm. Only middle names with at least three letters are retained.*

The output files from the cleaning steps are written out to tab-delimited text files and then processed further by a C++ program to link two files.

B. Stage 2: C++ Processing

Each dyad is processed by the C++ program in a slightly different manner, but following generally the same structure. For each pass over the data, blocking variables are created, data are sorted by block, and match attempts are made by comparing a single observation in one dataset, with a subset of observations in the other dataset (as defined by the blocking variables) and repeatedly calling the Jaro-Winkler algorithm to calculate a similarity score based on name. The link with the highest score is kept, and if that score is greater than or equal to the threshold for that block, then the link is saved. As the algorithm moves forward to other blocks, observations that have already been successfully linked are removed from the pool (though there are dyads where an observation in one dataset can be linked to multiple observations in another). This process continues, widening the blocks on each successive pass.

Details for the processing of each dyad are provided below.

1. EOUSA Matters OUT (Investigations Concluded)/USMS IN (Arrests)

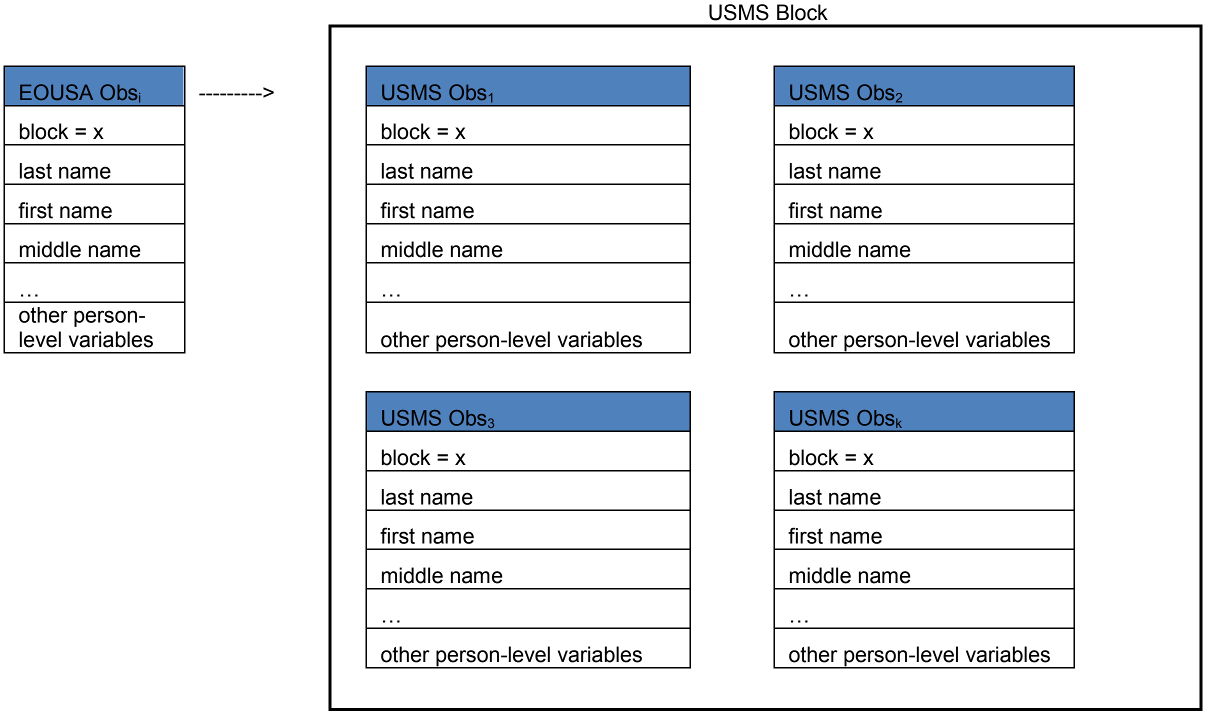
Initially, the EOUSA Matters IN file was used for this dyad. The Matters IN file however, suffers from several data limitations. Most notably, due to posting lags, not all observations will appear in the appropriate years' Matters IN SAF. However, a corresponding observation will appear in the Matters OUT SAF. For example, if a matter is received by EOUSA in 1999, but not posted in the data until 2000, then it will not appear in the 1999 Matters IN SAF (because it is not in the raw data as received by UI for 1999). However, if the same matter was also posted as a matter concluded in 2000, then it would appear in the 2000 Matters OUT SAF with a received date in 1999 and a closed date in 2000. These orphaned records will then only appear in the Matters OUT SAF and never appear in the Matters IN SAF. Additionally, for many observations the name field is richer[5] when using the Matters OUT (as opposed to the Matters IN) SAF. Further, for many observations in the Matters OUT file, the value of COURTNBR (a key blocking variable) has been assigned (for those matters that became cases filed in U.S. district court within the same fiscal year, as reflected in information posted in the same extract), whereas in the Matters IN file, the value of COURTNBR may not have been assigned yet..

At the start of the process, the entire standardized FY1994-2009 USMS IN file created in step 1 is read into memory, and blocking variables for the first pass are calculated. Then, the USMS observations are sorted by the first set of blocking variables. Next, observations from the standardized FY1994-2009 EOUSA file are read one-by-one and the blocking variable is calculated for a single observation. The EOUSA observation is then compared to all USMS observations in that block (see Figure V-1, below).

[5] That is, middle name fields are more populated in the OUT record and hyphenated names are more complete (a name such as Jose Hernandez in the IN might, for example, be recorded as Jose Hernandez-Gutierrez in the OUT).

Figure V-1 USMS IN and EOUSA Matters OUT Linking

Read in all USMS observations and process EOUSA MATTERS OUT one-by-one. Find links to EOUSA in the USMS data.



Compare EOUSA Obs$_i$ with USMS Obs$_1$ - USMS Obs$_k$ and keep the best match (USMS Obs$_j$). If the Jaro-Winkler score (based on name similarity) is above the threshold then EOUSA Obs$_i$ and USMS Obs$_j$ are considered a match

If the best match has a score that meets the threshold for the block, then a link is established and saved; otherwise, we make further attempts to match this EOUSA observation with the USMS observations in this block by permuting the EOUSA name under consideration.[6] If a match is still not found in this block, then we move on to the next EOUSA observation.

After processing all EOUSA observations, the USMS records are revisited, the second set of blocking variables are calculated and the USMS file is resorted. The EOUSA observations are reprocessed, and unlinked observations are examined one-by-one. The block for the EOUSA observation is calculated and compared to all USMS observations with corresponding blocking values. A single USMS observation may link to multiple EOUSA observations[7].

Table V-4 below shows all variables used in each block and the threshold used in determining if names are similar enough.

[6] Because the Jaro-Winkler distance is sensitive to ordering particularly at the front of the string, we must be careful in constructing our name strings for comparison. We hold the USMS name constant and, for example, use "LASTFIRSTMIDDLE" as the EOUSA name in the first try, and "FIRSTMIDDLELAST" in the next.

[7] If the name on the USMS file is JOHN or JANE DOE then it is excluded from blocks 2 – 4. Further, blocks 1 and 2 are the only block that will allow a USMS record to be linked multiple times to an EOUSA record.

Table V-4 Blocking Variables and Thresholds Used in USMS and EOUSA Linking

	Variable Description	USMS IN	EOUSA Matters OUT	Threshold
block 1	Docket number	CRTCNUM	COURTNBR	0.89
block 2	Arrest/Receive Year Month, District	ARDATE, DIST	ARREST_YM, RCV_YM, DISTRICT	0.93
block 3*	Arrest/Receive Year Month, First three letters of last name	ARDATE, first three letters of last name	ARREST_YM, RCV_YM, first three letters of last name	0.96
block 4	Arrest/Receive Year Month, First three letters of first name	ARDATE, first three letters of first name	ARREST_YM, RCV_YM, first three letters of first name	0.96

**NOTE: In this block only, if a match is not initially found, the year used on the EOUSA record is adjusted backward one year if the month is January – June and forward one year if the month is July – December to try and account for lags.*

DETAILED EXAMPLE

Consider the following example. Suppose an EOUSA observation is being compared to three USMS observations in the first block.

Table V-5 Detailed Example Applying the Jaro-Winkler Algorithm

	Raw Name	Cleaned Name	Name for Jaro-Winkler
EOUSA Obs	John Smith Jones	JOHN SMITH JONES	JOHNSMITHJONES
USMS Obs	James Jones	JAMES JONES	JAMESJONES
	John Smyth-Jones	JOHN SMYTHJONES	JOHNSMYTHJONES
	Jane Johnson	JANE JOHNSON	JANEJOHNSON

Just by looking at the records, we can see that the second USMS observation is the best match for the EOUSA observation. The Jaro-Winkler algorithm will produce the same result, and also decide if this match is “good enough”.

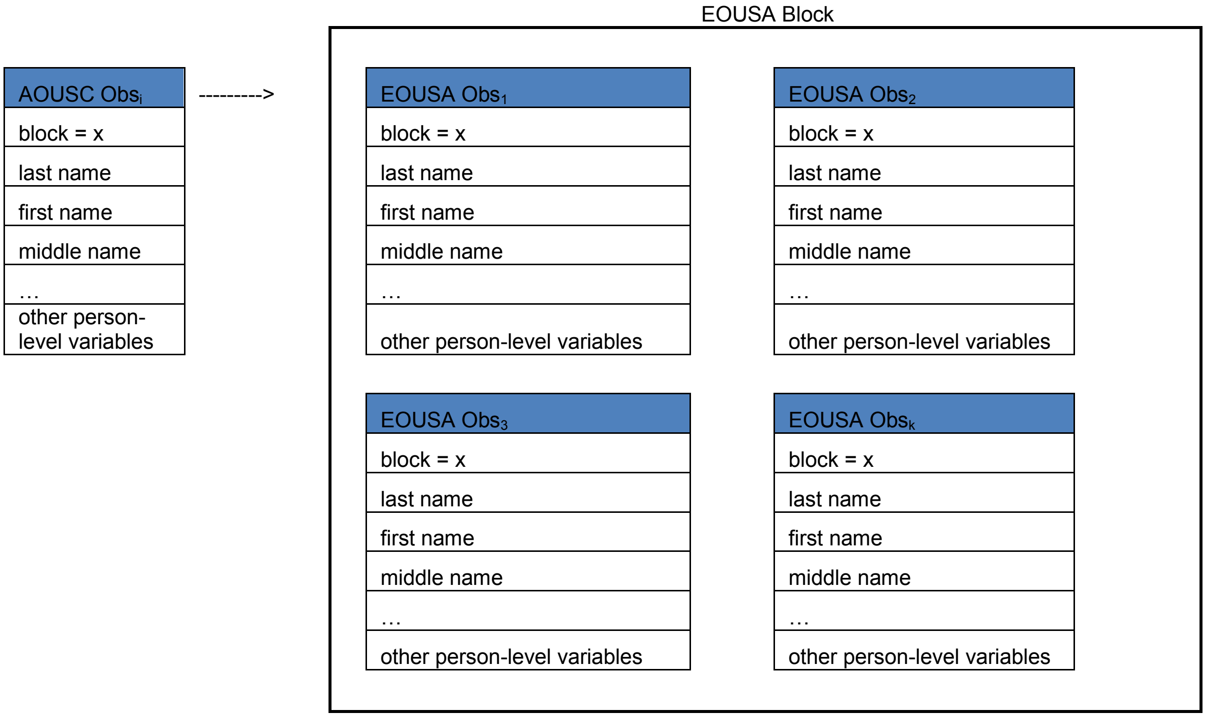
First, the EOUSA name “JOHNSMITHJONES” is compared to the first USMS observation “JAMESJONES”. This results in a Jaro-Winkler score of 0.6595. Next, we compare “JOHNSMITHJONES” and “JOHNSMYTHJONES”, resulting in a score of 0.98. Since 0.98 is greater than 0.6595, the link to the second observation is kept as the best match so far. We still go on to compare with the third observation, “JANEJOHNSON”, and this results in a score of 0.6179. Since this is not better than 0.98, the second observation is still the best match for the block. Finally, we compare 0.98 to the threshold for the block, 0.89. Since 0.98 ≥ 0.89, the link is saved and we consider these two observations to be matched.

2. AOUSC (Defendant-Cases)/EOUSA (Defendant-Cases)

For both the EOUSA OUT/AOUSC OUT link and EOUSA IN/AOUSC IN link, the entire standardized FY1994-2009 EOUSA file created in stage 1 is read into memory and blocking variables for the first pass are calculated. Next, observations from the standardized FY1994-2009 AOUSC file are read one-by-one and the blocking variable is calculated for a single observation. Then, the AOUSC observation is compared to all EOUSA observations in that block.

Figure V-2 AOUSC and EOUSA Linking

Read in all EOUSA IN/OUT observations and process AOUSC IN/OUT one-by-one. Finding links to AOUSC in the EOUSA data.



Compare AOUSC Obs$_i$ with EOUSA Obs$_1$ - EOUSA Obs$_k$ and keep the best match (EOUSA Obs$_j$). If the Jaro-Winkler score (based on name similarity) is above the threshold then AOUSC Obs$_i$ and EOUSA Obs$_j$ are considered a match

If the best match in this block has a score that meets the threshold for the block, then a link is established and saved; otherwise, we make further attempts to match this AOUSC record with the EOUSA records in the block by permuting the AOUSC name under consideration. If a link is not found in this block, then we move to the next AOUSC observation.

After processing all AOUSC observations for the first block, new EOUSA blocking variables are calculated and resorted. The AOUSC observations are then reprocessed one-by-one and unlinked observations are compared to an EOUSA block using the new set of blocking variables.

Tables V-6 and V-7 below show all variables used and the thresholds associated with each block for this dyad[8].

[8] Note that the only differences in blocking for the IN and OUT cohorts are the usage of MAGFLAG in the OUT cohort and the dates used.

Table V-6 Blocking Variables and Thresholds Used in AOUSC IN/EOUSA Cases IN Linking

	Variable Description	AOUSC IN	EOUSA IN	Threshold
block 1	Docket, Defendant Number	DCKT_YR, DCKT_NUM, DEFEND	COURTNBR, DEFNO4C	0.89
block 2	File Year Month, District	FIL_YY, FIL_MM, TRM_YM	FIL_YM, DISTRICT	0.93
block 3	File Year Month, First three letters of last name	FIL_YY, first three letters of last name	FIL_YM (year only),first three letters of last name	0.96
block 4	File Year Month, First three letters of first name	FIL_YY first three letters of first name	FIL_YM (year only), first three letters of first name	0.96

NOTES: No SEALED records processed in any block

Table V-7 Blocking Variables and Thresholds Used in AOUSC OUT/EOUSA Cases OUT and Matters OUT Linking

	Variable Description	AOUSC OUT	EOUSC Cases OUT and Matters OUT	Threshold
block 1	Docket, Defendant Number, Magistrate flag	DCKT_YR, DCKT_NUM, DEFEND, FILEMAG	COURTNBR, DEFNO4C, MAGFLAG	0.89
block 2	One of Disposition Year Month, Termination Year Month, File Year Month, District, Magistrate flag	DISP_YY, DISP_MM, FIL_YY, FIL_MM, TRM_YM, FILEMAG	DISP_YM, FIL_YM, TERM_YM, DISTRICT, MAGFLAG	0.93
block 3	One of Disposition Year, Termination Year, File Year, First three letters of last name	DISP_YY, FIL_YY, TRM_YM (year only), first three letters of last name	DISP_YM (year only), FIL_YM (year only), TERM_YM (year only), first three letters of last name	0.96
block 4	One of Disposition Year, Termination Year, File Year, First three letters of first name	DISP_YY, FIL_YY, TRM_YM (year only), first three letters of first name	DISP_YM (year only), FIL_YM (year only), TERM_YM (year only), first three letters of first name	0.96

NOTES: MAGFLAG for EOUSA is set equal to 1 if the observation from the EOUSA Matters OUT SAF

Date determined as follows: if magistrate case then dispyearmonth is used, else termyearmonth is used if not blank, otherwise fileyearmonth is used

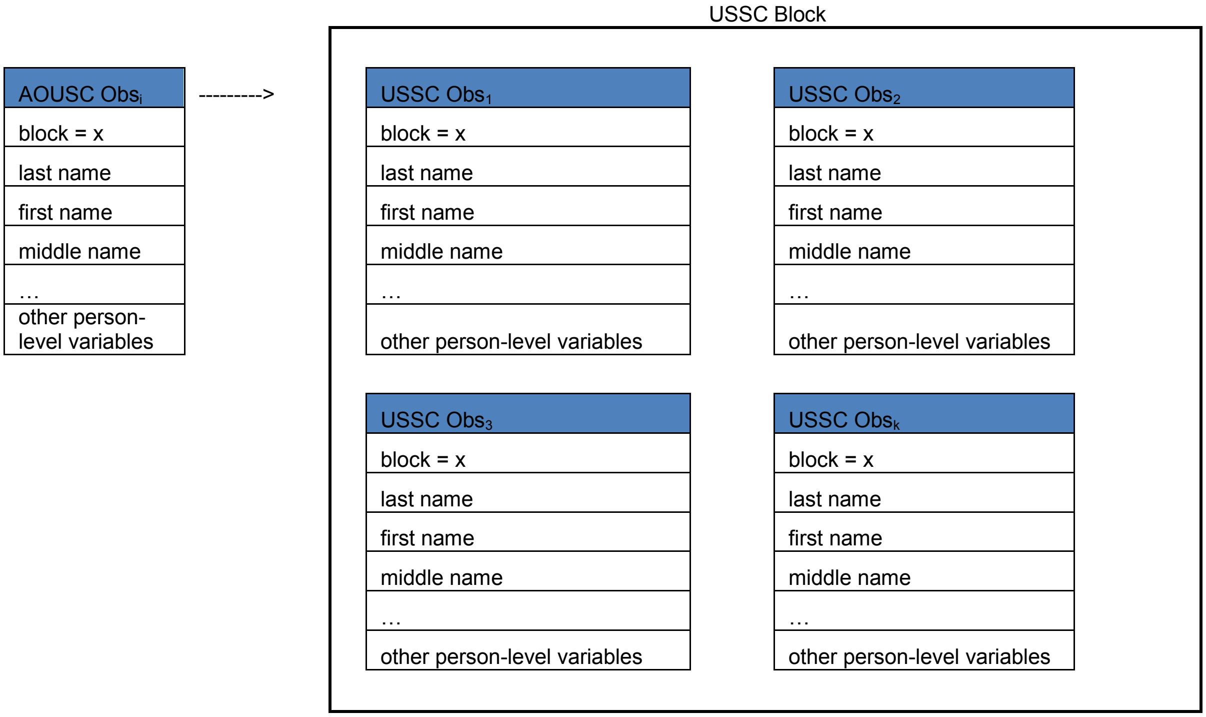
No SEALED records processed in any block

3. AOUSC OUT (Defendant-Cases)/USSC OUT (Offenders Sentenced)

At the beginning of the process, the entire standardized FY1994-2009 USSC file created in stage 1 is read into memory and blocking variables for the first pass are calculated. Then the USSC observations are sorted by the first set of blocking variables. Next, observations from the standardized FY1994-2009 AOUSC file are read one-by-one and the blocking variable is calculated for a single observation. Then, the AOUSC observation is compared to all USSC observations in that block.

Figure V-3 AOUSC OUT and USSC OUT Linking

Read in all USSC OUT observations and process AOUSC OUT one-by-one. Finding links to AOUSC in the USSC data.



Compare AOUSC Obs$_i$ with USSC Obs$_1$ - USSC Obs$_k$ and keep the best match (USSC Obs$_j$). If the Jaro-Winkler score (based on name similarity) is above the threshold then AOUSC Obs$_i$ and USSC Obs$_j$ are considered a match

If the best match in this block has a score that meets the threshold for the block, then a link is established and saved; otherwise, we make further attempts to match this AOUSC observation with the USSC observations in this block by permuting the AOUSC name under consideration. If a match is not found in this block, then we move on to the next AOUSC observation.

After processing all AOUSC observations, the USSC records are revisited and the second set of blocking variables is calculated and the USSC file is resorted. The AOUSC observations are then reprocessed, and unlinked observations are once again examined one-by-one. The block for the AOUSC observation is calculated and once again is compared to all USSC observations with corresponding blocking values.

Table V-8 below shows all variables used in each block and the thresholds used in determining if names are similar enough for this dyad.

Table V-8 Blocking Variables and Thresholds Used in AOUSC OUT/EOUSA Cases OUT and Matters OUT Linking

	Variable Description	AOUSC OUT	USSC	Threshold
block 1	District, Judge, Docket Year, Docket Number	DISTRICT, TRMJUDGE1, TRMJUDGE2, TRMJUDGE3, TRMJUDGE4, DCKT_YR, DCKT_NUM	DISTRICT, JUDGE, DOCKETID	0.8
block 2	Sentence Date, District	SENT_YY, SENT_MM, DISTRICT	SENTDATE, DISTRICT	0.93
block 3	Sentence Date, First three letters of last name	SENT_YY, SENT_MM, first three letters of last name	SENTDATE, first three letters of last name	0.96

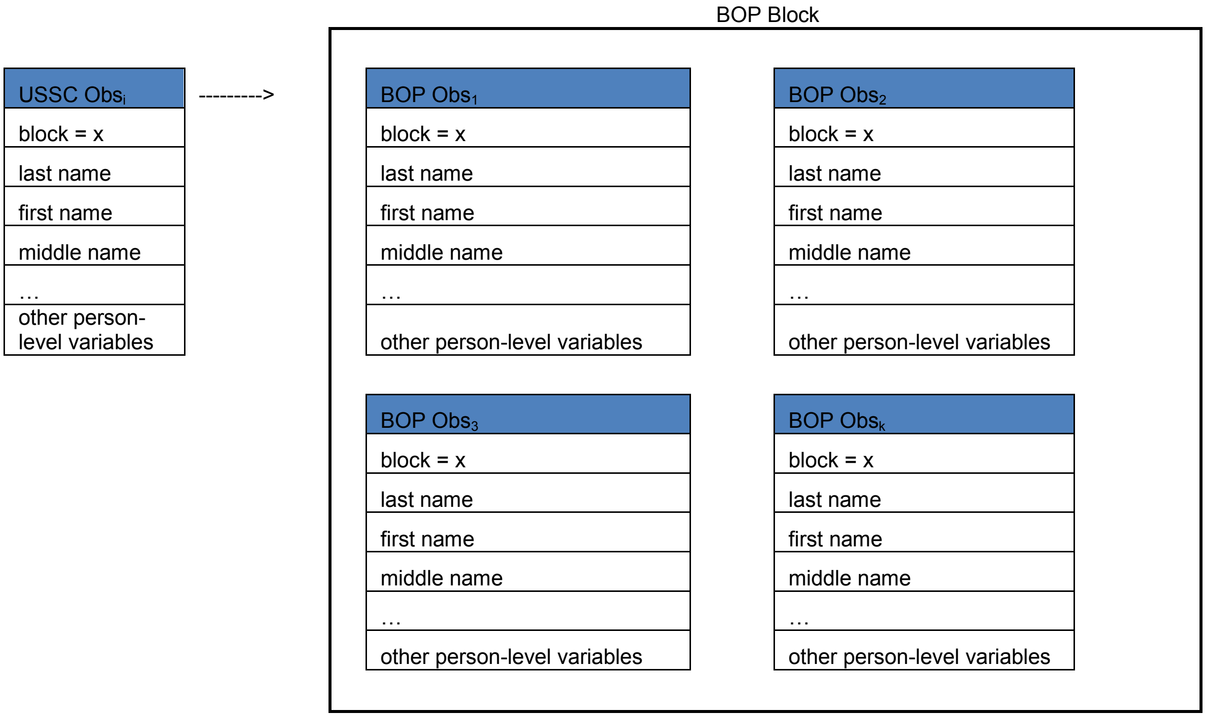
NOTES: No SEALED records processed in any block

4. BOP IN (Entering Prisoners)/USSC OUT (Offenders Sentenced to Prison)

At the beginning of the process, the entire standardized FY1994-2009 BOP file created in stage 1 is read into memory and blocking variables for the first pass are calculated. Then, the BOP observations are sorted by the first set of blocking variables. Next, observations from the standardized FY1994-2009 USSC file are read one-by-one and the blocking variable is calculated for a single observation. The USSC observation is then compared to all BOP observations in that block.

Figure V-4 BOP IN and USSC OUT Linking

Read in all BOP IN observations and process USSC OUT one-by-one. Finding links to USSC in the BOP data.



Compare USSC Obs$_i$ with BOP Obs$_1$ - BOP Obs$_k$ and keep the best match (BOP Obs$_j$). If the Jaro-Winkler score (based on name similarity) is above the threshold then USSC Obs$_i$ and BOP Obs$_j$ are considered a match

If the best match in this block has a score that meets the threshold for the block, then a link is established and saved; otherwise, we make further attempts to match this USSC observation with the BOP observations in this block by permuting the USSC name under consideration. If a match is not found in this block, then we move on to the next USSC observation.

After processing all USSC observations, the BOP records are revisited and the second set of blocking variables is calculated and the BOP file is resorted. The USSC observations are then reprocessed, and unlinked observations are once again examined one-by-one. The block for the USSC observation is calculated and once again is compared to all BOP observations with corresponding blocking values. A single BOP observation may link to multiple USSC observations.

In the final three loops, we do not call the Jaro-Winkler algorithm. Instead, we assume that Marshals Number, FBI Number and Social Security Number have been stored accurately by both agencies and a match is made if values match in both the BOP IN and the USSC files. If the number is missing or coded as unknown, then the observation is excluded from the block. Matches made using Marshals Number, FBI Number or Social Security Number will have the Jaro-Winkler score set to 99.

Table V-9 below shows all variables used in each block and the thresholds used in determining if names are similar enough to be considered a match for this dyad.

Table V-9 Blocking Variables and Thresholds Used in BOP IN/ USSC OUT Linking

	Variable Description	BOP IN	USSC OUT	Threshold
block 1	Sentence Date, Docket Number	SENTDT, DOCKTNO	SENTDATE, DOCKETID	0.89
block 2	Sentence Date, FBI Number	SENTDT, FBINUM	SENTDATE, FBINUM	NA
block 3	Sentence Date, Marshals Number	SENTDT, REGNO	SENTDATE, MARSLNUM	NA
block 4	Sentence Date, Social Security Number	SENTDT, SSNNUM	SENTDATE, DEFSSN	NA

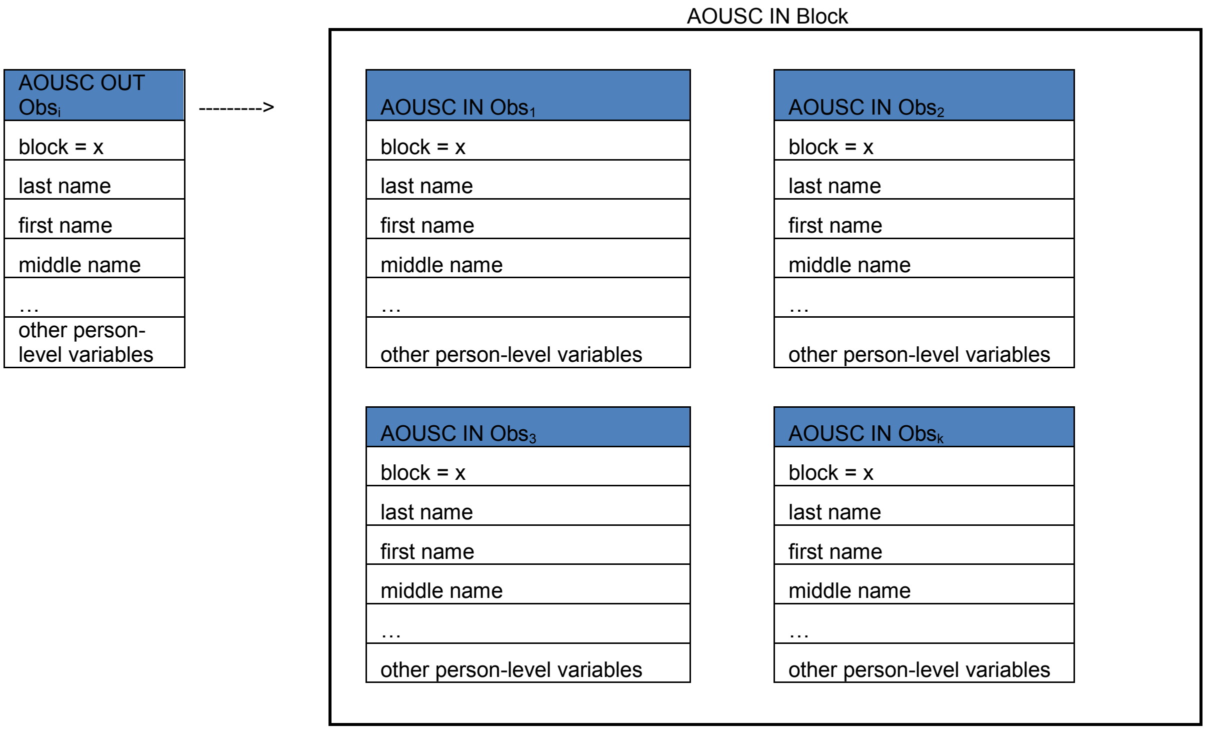
NOTES: For blocks 2,3 and 4, observations where the number is missing, or coded as unknown (e.g. a SSN of 999-99-9999) are not included

5. AOUSC INTRA-AGENCY LINKS

At the beginning of the process, the entire standardized FY1994-2009 AOUSC IN file created in stage 1 is read into memory and blocking variables for the first pass are calculated. Then the AOUSC IN observations are sorted by the first, and in this case only, set of blocking variables. Next, observations from the standardized FY1994-2009 AOUSC OUT file are read one-by-one and the blocking variable is calculated for a single observation. Then, the AOUSC OUT observation is compared to all AOUSC IN observations in that block.

Figure V-5 AOUSC IN and AOUSC OUT Linking

Read in all AOUSC IN observations and process AOUSC OUT one-by-one. Finding links to AOUSC OUT in the AOUSC IN data.



Compare AOUSC OUT Obs$_i$ with AOUSC IN Obs$_1$ - AOUSC IN Obs$_k$ and keep the best match (AOUSC IN Obs$_j$). If the Jaro-Winkler score (based on name similarity) is above the threshold then AOUSC OUT Obs$_i$ and AOUSC IN Obs$_j$ are considered a match

If the best match in this block has a score that meets the threshold for the block, then a link is established and saved; otherwise, we make further attempts to match this AOUSC OUT observation with the AOUSC IN observations in this block by permuting the AOUSC OUT name under consideration. If a match is not found in this block, then we move on to the next AOUSC OUT observation.

Table V-10 below shows all variables used and the thresholds used in determining if names are similar enough.

Table V-10 Blocking Variables and Thresholds Used in AOUSC IN/AOUSC OUT Linking

	Variable Description	**AO Variables**	**Threshold**
block 1	District, Docket Year, Docket Number, Defendant Number	DISTRICT, DCKET_YR, DCKT_NUM, DEFEND	0.89

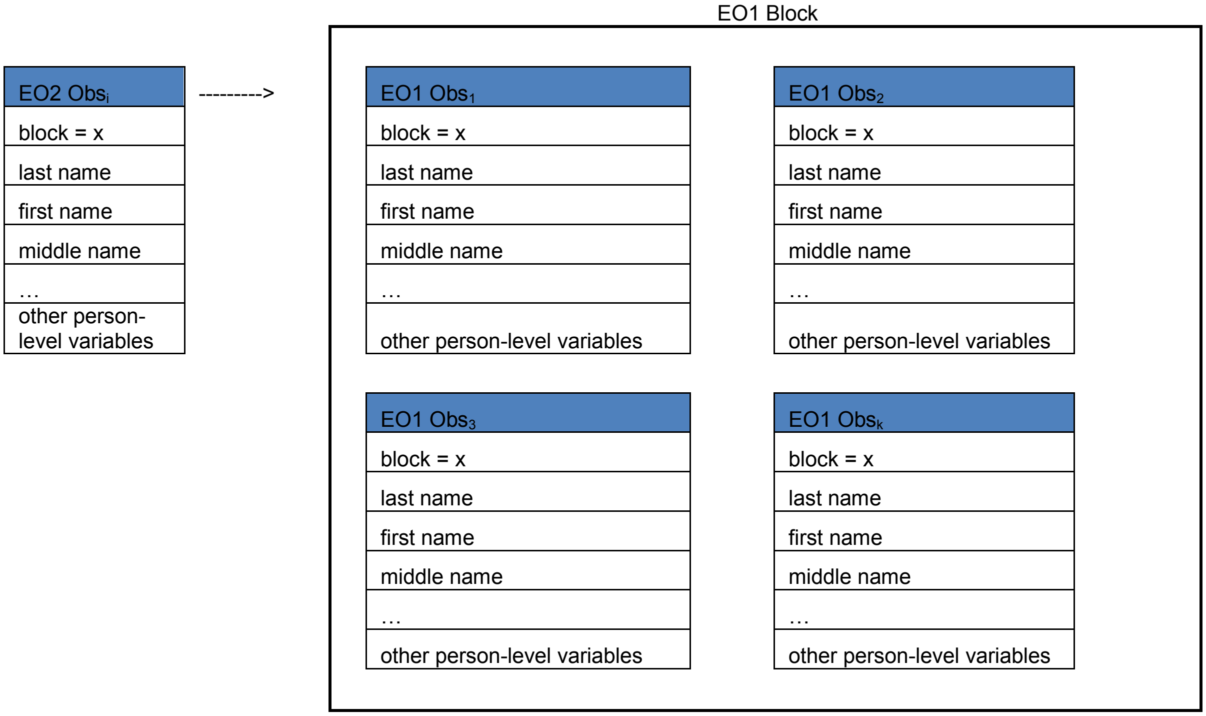
NOTES: No SEALED records processed

6. EOUSA INTRA-AGENCY LINKS

At the beginning of the process, one of the entire standardized FY1994-2009 EOUSA files created in stage 1 is read into memory and blocking variables for the first pass are calculated. Then the observations are sorted by the first set of blocking variables. Next, observations from the second standardized FY1994-2009 EOUSA file are read one-by-one and the blocking variable is calculated for a single observation. Finally, the EOUSA observation from the second data set is compared to all EOUSA observations from the first dataset in that block.

Figure V-6 EOSUA Intra Links

Read in all EO1 observations and process EO2 one-by-one. Finding links to EO2 in the EO1 data.



Compare EO2 Obs$_i$ with EO1 Obs$_1$ - EO1Obs$_k$ and keep the best match (EO1Obs$_j$). If the Jaro-Winkler score (based on name similarity) is above the threshold then EO2 Obs$_i$ and EO1 Obs$_j$ are considered a match

If the best match in this block has a score that meets the threshold for the block, then a link is established and saved; otherwise, we make further attempts to match this EO2 observation with the EO1 observations in this block by permuting the EO2 name under consideration. If a match is not found in this block, then we move on to the next EO2 observation.

After processing all EO2 observations, the EO1 records are revisited, the second set of blocking variables is calculated and the EO1 file is resorted. The EO2 observations are then reprocessed, and unlinked observations are once again examined one-by-one. The block for the EO2 observation is calculated and once again is compared to all EO1 observations with corresponding blocking values.

There are four sets of dyads that make up the set of EOUSA Intra-Agency links, shown below in Table V-11 and Table V-12.

Table V-11 List of EOUSA Intra Agency Links

EO1	EO2
Cases IN	Cases OUT
Matters OUT	Cases OUT
Matters OUT	Cases IN
Matters IN	Matters OUT

Each of these pairs has a slightly different set of blocking variables as follows:

Table V-12 Blocking Variables and Thresholds Used in EOUSA Intra-Agency Links

		Variable Description	EOUSA Variables	Threshold
Cases IN/Cases OUT	block 1	District, LIONS Number	DISTRICT, LIONS	0.89
	block 2	File Year Month or Termination Year Month, District	FILEYM, TERMYM, DISTRICT	0.93
	block 3	File Year Month or Termination Year Month, Last name	FILEYM, TERMYM, first three letters of last name	0.93
Matters OUT/Cases OUT	block 1	District, LIONS Number	DISTRICT, LIONS	0.89
	block 2	Disposition Year Month or File Year Month or Termination Year Month, District	DISPYM, FILEYM, TERMYM, DISTRICT	0.93
	block 3	Disposition Year Month or File Year Month or Term Year Month, Last name	DISPYM, FILEYM, TERMYM, first three letters of last name	0.93
Matters OUT/Cases IN	block 1	District, LIONS Number	DISTRICT, LIONS	0.89
	block 2	Disposition Year Month or File Year Month or Termination Year Month, District	DISPYM, FILEYM, TERMYM, DISTRICT	0.93
	block 3	Disposition Year Month or File Year Month or Term Year Month, Last name	DISPYM, FILEYM, TERMYM, first three letters of last name	0.93
Matters IN/Matters OUT	block 1	District, LIONS Number	DISTRICT, LIONS	0.89
	block 2	Disposition Year Month, District	DISPYM, DISTRICT	0.93
	block 3	Disposition Year Month, Last name	DISPYM, first three letters of last name	0.93

VI. Results of Dyad Linking

This section provides information on the extent of links found across the dyads. The table below shows the percent of observations in a particular year's SAF that are linked to another observation in the associated dyad with any screening criteria accounted for prior to calculating the rate. For example, 72.95% of AOUSC observations where the variable OUTCOME indicates a conviction in the 1994 AOUSC OUT SAF have a link to a USSC OUT observation. Note that this USSC observation can be in any year 1994-2009.

Table VI-1 Results of Inter-Agency Linking

	EOUSA Matters OUT - USMS IN		**EOUSA IN - AOUSC IN**		**EOUSA OUT - AOUSC OUT**		**AOUSC OUT - USSC OUT**		**BOP IN - USSC OUT**	
	% EOUSA linked	**% USMS linked**	**% AOUSC linked**	**% EOUSA linked**	**% AOUSC linked**	**% EOUSA linked**	**% AOUSC linked**	**% USSC linked**	**% BOP linked**	**% USSC linked**
1994	47.26%	69.95%	74.25%	76.51%	75.07%	68.74%	72.95%	91.24%	67.01%	88.13%
1995	50.78%	74.71%	79.27%	78.23%	78.95%	66.88%	72.67%	89.71%	78.75%	88.01%
1996	53.25%	73.76%	78.51%	79.26%	80.26%	72.74%	74.69%	92.12%	81.32%	87.22%
1997	56.47%	74.28%	75.16%	75.47%	80.60%	70.76%	80.10%	92.76%	84.83%	86.08%
1998	60.26%	73.71%	76.97%	87.38%	78.92%	68.72%	75.83%	91.08%	82.50%	85.84%
1999	60.31%	75.11%	79.24%	87.67%	80.21%	68.93%	77.08%	91.65%	80.72%	84.35%
2000	59.86%	73.39%	80.95%	86.90%	83.46%	68.93%	80.33%	91.48%	82.49%	84.25%
2001	59.91%	73.69%	81.03%	87.07%	83.55%	68.25%	80.26%	91.83%	82.07%	85.27%
2002	60.38%	75.41%	80.40%	86.49%	85.79%	70.39%	82.82%	92.38%	86.62%	83.72%
2003	60.27%	76.21%	78.87%	87.98%	84.29%	70.23%	85.48%	92.29%	89.51%	82.86%
2004	62.59%	77.21%	80.10%	80.99%	84.24%	62.99%	85.91%	91.69%	87.64%	82.59%
2005	63.18%	77.77%	79.04%	76.86%	83.19%	63.70%	84.97%	91.52%	87.09%	82.60%
2006	64.83%	74.41%	80.22%	76.65%	84.36%	62.55%	82.99%	91.36%	84.41%	82.71%
2007	66.88%	73.77%	80.31%	79.33%	82.76%	61.86%	85.38%	92.98%	85.04%	81.37%
2008	73.69%	89.13%	82.64%	80.56%	84.92%	49.26%	86.44%	93.62%	86.57%	78.23%
2009	72.41%	88.72%	81.60%	79.98%	85.92%	49.47%	87.51%	93.53%	86.86%	59.85%

Some observations regarding these link rates are worth noting:

- First, although manual checks of the accuracy of the links have been conducted on a portion of the matches; false positive and negative hits are still present. Hence, a match rate does not reflect absolute accuracy. Instead it simply means that the links were assessed and further improvements could not be made. Some dyads are clearly weaker than others; the link between AOUSC and EOUSA IN for example, is weaker than the link between AOUSC OUT and USSC.

- The EOUSA link rate in the EOUSA Matters OUT/USMS IN dyad is low because there are many EOUSA observations with a disposition indicating a criminal declination. It is expected that most of these observations are not in the USMS data as the matter never resulted in an arrest. However, a portion of these records[9], do have links to the USMS IN SAF.

- Also, in the case of the AOUSC OUT/ EOUSA OUT link, there was a large increase in magistrate matters in fiscal year 2008 when compared to 2007.[10] Additionally, appeals cases are in EOUSA cases out, but not AOUSC. Removing these observations in EOUSA results in a link rate of 81.08% for EOUSA in 2003. Further, EOUSA has a few duplicate records (<1%) where sentencing or disposition information has been updated that need to be investigated further. The AOUSC data contains misdemeanors that EOUSA does

[9] 20% of EOUSA Matters OUT records with a DISPOS=”DE” in FY2008 have a link to a USMS observation.

[10] This is not unexpected; there was a change in DOJ policy that was implemented that involved prosecuting minor immigration petty offenses/infractions in the SW districts that in the past would not have been handled in federal criminal court. The 2008 data were examined by district to verify that the increase comes from districts in the southwest.

not. The AOUSC data contains some persons charged with class B and C misdemeanors who were proceeded against before U.S. district court judges, whereas the EOUSA data does not. Additionally, there was a major change in the EOUSA data structure beginning with FY 2004. It appears that the link rate is lower for years with this new structure. Further investigation is needed in this area; it appears that it is likely due to an increase in magistrate matters.

- The percent of AOUSC observations with links to USSC is slightly lower in part because USSC does not contain information pertaining to juveniles who were adjudicated delinquent.

- The BOP IN/USSC OUT link rate in 2009 is low partly because a record is not immediately entered into the BOP data upon sentencing. That is, a record is present in the USSC data, but not yet in the BOP data. When the dyad linking was run for years 1994 – 2008, the percent of unlinked USSC observations in 2008 was about 40%; when 2009 data was added and the dyad linking re-run, the percent of unlinked USSC observations in 2008 dropped to 21.77%. This pattern has been observed over several iterations of adding new data, and in each case, adding the next year's data improves the link rates for USSC in the year prior.

- The residual unlinked observations for each dyad have been examined in detail. We have looked for any patterns in the unlinked records to determine if additional screening should be considered. The analysis of AOUSC and EOUSA seems to indicate that the unlinked AOUSC observations are more likely to be misdemeanor offenses in both the IN and OUT cohorts. This may mean that additional screening rules should be developed for these dyads.

Detailed results for each dyad can be found in the attached appendix.

Table VI-2 Results of Intra-Agency Linking

	AOUSC Cases IN - AOUSC Cases OUT		**EOUSA Cases IN- EOUSA Cases OUT**		**EOUSA Matters OUT - EOUSA Cases IN**		**EOUSA Matters OUT - EOUSA Cases OUT**		**EOUSA Matters OUT - EOUSA Matters IN**	
	% AOUSC Cases IN	**% AOUSC Cases OUT**	**% EOUSA Cases IN**	**% EOUSA Cases OUT**	**% EOUSA Matters OUT**	**% EOUSA Cases IN**	**% EOUSA Matters OUT**	**% EOUSA Cases OUT**	**% EOUSA Matters OUT**	**% EOUSA Matters IN**
1994	92.37%	48.31%	92.20%	38.81%	51.83%	84.97%	47.20%	37.50%	57.79%	87.46%
1995	94.31%	86.10%	89.83%	86.11%	54.02%	86.32%	44.38%	75.16%	79.36%	81.25%
1996	92.14%	93.53%	86.72%	94.69%	57.55%	86.18%	40.79%	82.30%	87.77%	71.78%
1997	91.94%	95.05%	86.74%	89.95%	60.02%	84.24%	54.96%	65.15%	72.67%	91.00%
1998	93.34%	95.59%	92.17%	89.03%	61.27%	93.79%	56.50%	74.03%	86.03%	92.98%
1999	93.28%	95.52%	93.50%	86.34%	59.83%	94.21%	56.24%	78.80%	89.37%	93.46%
2000	92.70%	96.12%	92.71%	86.76%	61.84%	94.01%	57.53%	80.40%	92.62%	93.48%
2001	93.12%	96.78%	92.82%	86.48%	60.61%	93.90%	56.42%	79.89%	93.20%	92.90%
2002	91.64%	96.59%	92.74%	85.75%	61.02%	94.25%	57.40%	80.96%	93.32%	93.71%
2003	90.01%	95.45%	89.27%	86.57%	61.05%	96.80%	56.64%	81.28%	93.53%	62.84%
2004	89.36%	94.33%	91.16%	89.67%	57.24%	90.49%	53.54%	87.11%	79.00%	95.30%
2005	87.61%	92.48%	90.56%	94.35%	57.29%	87.75%	53.17%	87.46%	86.21%	94.06%
2006	87.57%	90.44%	89.73%	94.48%	56.33%	87.46%	51.91%	85.17%	88.60%	92.53%
2007	86.18%	90.41%	87.36%	94.43%	55.67%	89.46%	50.00%	85.72%	89.70%	90.47%
2008	79.86%	90.87%	79.39%	95.29%	46.51%	90.24%	38.04%	87.83%	92.48%	88.84%
2009	41.05%	91.61%	40.66%	95.98%	45.20%	90.51%	17.54%	89.17%	92.65%	79.85%

The above table, Table VI-2, shows the results of the intra-agency dyad links. For example, 91.94% of the AOUSC Cases IN observations in 1997 were linked to an AOUSC Cases OUT observation. Note that, as with the other dyads, these AOUSC OUT links could be from any year.

- In general, the intra-agency links perform well. In the case of the AOUSC links, there is a decrease in the number of links at the tails as we would expect – a decrease in the percent of AOUSC Cases OUT links in the early period, and a decrease in the percent of the AOUSC Cases IN links in the later period.

- As for the EOUSA intra links, there are several points worth noting. Not all EOUSA Matters become cases, thus the lower match rates for EOUSA Matters OUT when linking to EOUSA Cases IN and EOUSA Cases OUT. Additionally, there is a data issue with the 2003 EOUSA Matters IN file. There are a significant number of records on the 2003 Matters IN file that have blank and/or missing data on key blocking and matching variables. These observations are likely matched to the 2004 Matters OUT file (which has a corresponding dip in link rate). Because there are in most cases two defendants (who can be listed in different orders on the 2003 Matters IN file and the 2004 Matters OUT file), though, it is not possible to identify which observations should be linked.

As with the inter-agency links, detailed results for each intra-agency dyad can be found in the attached appendix.

A. Comparison to First Generation System

For each dyad, the results of the paired-agency dyad system have been compared to the linking results obtained using the first generation system. In general, the new methodology results in more links per dyad than the first generation system. For dyads where screening rules are used, caution should be taken when comparing results. Because the first generation system did not have such screening rules, there will be cases where there was a link in the first generation system but not in the dyad system (because one or both observations did not meet the screening conditions). Further, in the later years, an accurate comparison is not possible because at the time of this analysis the results from the first generation system only contained links for fiscal years 1994-2005 (resulting in links that are made in the dyad system that are impossible to make in the first generation system). That being said, for most dyads, the links made in both systems are similar. The dyad with the largest differences, the BOP IN/USSC OUT dyad, also has the most screening conditions and makes use of personal identifiers not used in the first generation system. The dyad with the most common links, AOUSC OUT/USSC OUT, is most similar to the methodology used in the first generation system.

Table VI-3 Percent of Dyad and First Generation System Links identical (observations from FY1994-2005)

Dyad	Percent
EOUSA Matters OUT/USMS IN	51.66
AOUSC IN/EOUSA IN	59.00
AOUSC OUT/EOUSA OUT	62.10
AOUSC OUT/USSC OUT	64.27
BOP IN/USSC OUT	36.76

More detailed comparisons between the two systems, by dyad, can be found in the Appendix.

B. Putting multiple dyads together

It is possible to put results from multiple dyads together, though in some cases, special considerations must be made as for some dyads, an identifier may be output more than once.

For example, if a user wanted to examine EOUSA OUT and AOUSC OUT links and also AOUSC OUT and USSC OUT links, he or she could take the EOUSA-AOUSC OUT results and merge them by AOSeqNum to the AOUSC-USSC results.

Table VI-4 Example Output and Merge

AOUSC-USSC output

aoYear	scYear	aoSeqNum	scSeqNum	jwScore
2006	2006	AOC20060123456	SCR20060987654	1

AOUSC-EOUSA OUT output

aoYear	eoYear	aoSeqNum	eoSeqNum	jwScore	magflag
2006	2006	AOC20060123456	LIO2007056789	1	0

Merged file

eoYear	aoYear	scYear	eoSeqNum	aoSeqNum	scSeqNum
2006	2006	2006	LIO2007056789	AOC20060123456	SCR20060987654

Using year, SeqNum (and magflag where necessary), the user can then merge variables from the SAFs and, as an example, be able to analyze demographic characteristics and sentencing information (from USSC) by arresting agency (from EOUSA).

Care should be taken when merging more than one dyad together. If, for instance, we were interested in the looking at EOUSA observations linked to BOP, since the path goes through USSC, the end result is those persons sentenced to prison. For example, of the 89,510 EO observations in 1999, 68.93% or 61,700 observations have a link in the AO data. Of those AO observations, only 55,875 are convicted and pass the screen to be eligible to be linked to USSC. There are 45,522 AO observations that are linked to USSC. Of those, there are only 38,209 USSC observations that are eligible to be linked to BOP. Finally, there were 33,164 USSC observations that were linked to BOP. Thus, only 33,164 / 89,510 = 37% of the EO observations in 1999 make it all the way through to the BOP links. Some of the EO records that did not trace all the way through because of case processing decisions, or exit points in the system (e.g., matters that were declined for prosecution, case acquittals or dismissals, and cases sentenced to probation rather than prison).

Table VI-5 EOUSA Cases Out and Magistrate Matters OUT Linked Through To USSC OUT

Year	EO obs	EO obs linked to AO out	% linked	AO OUT linked to EO that are convicted	USSC Obs linked to AO Out	% linked	USSC Obs linked to AO that are sentenced to prison	BOP In Obs that are linked to USSC	% linked
1994	68,796	47,289	68.74%	40,306	32,850	81.50%	25,913	23,266	89.79%
1995	68,640	45,907	66.88%	39,137	31,480	80.44%	25,080	22,485	89.65%
1996	69,241	50,365	72.74%	44,122	35,727	80.97%	29,237	26,015	88.98%
1997	74,855	52,967	70.76%	47,055	39,988	84.98%	32,879	28,852	87.75%
1998	80,933	55,616	68.72%	50,083	39,729	79.33%	33,409	29,389	87.97%
1999	89,510	61,700	68.93%	55,855	45,522	81.50%	38,209	33,164	86.80%
2000	93,434	64,403	68.93%	59,161	49,826	84.22%	42,483	36,621	86.20%
2001	95,074	64,892	68.25%	59,437	49,825	83.83%	42,566	37,008	86.94%
2002	98,821	69,554	70.38%	63,569	54,825	86.24%	47,094	40,095	85.14%
2003	103,238	72,507	70.23%	66,675	59,333	88.99%	51,333	43,143	84.05%
2004	113,289	71,360	62.99%	65,145	58,648	90.03%	51,040	42,780	83.82%
2005	114,365	72,851	63.70%	67,154	60,400	89.94%	53,233	44,285	83.19%
2006	119,618	74,824	62.55%	69,283	61,554	88.84%	54,448	45,526	83.61%
2007	118,994	73,611	61.86%	68,057	62,234	91.44%	55,017	45,182	82.12%
2008	158,957	78,300	49.26%	72,389	66,710	92.15%	59,354	46,672	78.63%
2009	166,815	82,528	49.47%	76,470	71,131	93.02%	63,708	38,505	60.44%

VII. Obtaining the Dyad Linking Files

The dyad linking files are available on a restricted use basis for download through the National Archive of Criminal Justice Data (http://www.icpsr.umich.edu/icpsrweb/content/NACJD/guides/fjsp.html). Persons interested in obtaining these files and the SAFs from NACJD must agree to abide by Federal laws and scientific standards regarding human subject protections.

The dyad linking files contain several variables that are essential to performing links: the randomly generated sequential id variables ("SeqNums") that are agency-specific, the SAF year variable, and the Jaro-Winkler score for each link. For example, the following is what an excerpt of output from the AOUSC-USSC output might look like:

Figure VII-1 Example Output From Dyad Linking AOUSC OUT and USSC OUT

```
aoYear  scYear  aoSeqNum          scSeqNum          jwScore
  1994    1994  AO0940000001      SCM950000002            1
  2003    2003  AOC2003000000017  SCR2003000000009     0.89
  2009       0  AOC2009090123456  NULL                   -1
```

The values for aoYear and scYear give the SAF year of the AOUSC and USSC observation listed (should the ids somehow be duplicated across years). The values aoSeqNum and scSeqNum list the UI created ids for the linked records, and jwScore gives the calculated Jaro-Winkler score for this link. When a link is not found, 0 is output for year, NULL for id and -1 for jwScore. A record with a value of scYear = 0, scSeqNum="NULL" and jwScore =-1 represents an AOUSC observation that went unlinked, and similarly, a record with a value of aoYear=0, aoSeqNum=”NULL” and jwScore=-1 represents a USSC observation that remained unlinked.

The output file formats for other dyads are identical with the exception of the output for AOUSC OUT/EOUSA OUT link. There is an additional variable, magflag, on this output file. This is a 0/1 indicator of whether the EOUSA observation came from the matters file (1 for yes, 0 for no).

VIII. Conclusion

The FJSP dyad linking system provides a method for tracking person-cases through the various stages of the federal criminal justice system. Further, it allows users to fill in missing information, such as demographics, for an agency cohort file with data from another agency. For example, demographics such as age and education are not included in the AOUSC criminal data but are available from the USSC data. By linking the USSC information to the AOUSC records, the resulting analytic file can be augmented with important demographic information that may be necessary for analysis. Other examples of analytic applications of the dyad link file include: calculating case-processing time and identifying case processing bottlenecks; examining case processing decisions by race and gender; and tracking a specific offense type from arrest to sentencing. There are countless others. However, users should also be aware that, for a minor proportion of the linked records, certain information may be different or conflict between the two agency sources when common data elements are recorded in both systems. In such situations, users must exercise their own judgments about which data source contains the more accurate and complete information.

In general, the dyad-based system improves upon the first generation system, adding more links and refining those links that are made through the use of screening conditions. As of the date of this publication, comprehensive analyses of unlinked records across the dyads have not yet been conducted. Should systematic differences exist between the sets of linked and unlinked records, any analyses performed on the set of linked records could produce results that are biased, and therefore may need to be addressed or corrected for (e.g., by applying statistical weighting techniques). Users of the dyad link system are encouraged to consider this fact when conducting and reporting on analyses that utilize the FJSP dyad link files.

For more information about obtaining FJSP SAFs and linking files, which are available on a restricted use basis, please consult the NACJD website, http://www.icpsr.umich.edu/icpsrweb/content/NACJD/guides/fjsp.html.

For guidance in using the linking files and for an example of its use please see the codebook available at the NACJD website, http://www.icpsr.umich.edu/icpsrweb/NACJD/studies/30701/documentation.

Appendices

A. *Jaro-Winkler Implementation*

The Jaro-Winkler distance is a variant of the Jaro distance (d_j).

$d_j = 1/3 (c/|s_1| + c/|s_2| + (c-t)/c)$

Where:

- s_1 and s_2 are the strings being compared.
- c is the number of common characters. A common character from s_1 is found in s_2 if it exists in s_2 in a position that is within a distance of half the length of the longer string from the position in s_1.
- t is the number of transpositions. The number of transpositions is calculated as the greatest integer of half of the number of out-of-order common character pairs. For example, when comparing the strings “PETER” and “PEETR”, there are two character pairs “TE” and “ET” that are out of order, resulting in one transposition.

The Jaro-Winkler distance improves upon the Jaro distance by increasing the score when up to the first four initial characters are common across strings. Thus,

$d_{jw} = d_j + L/10(1-d_j)$

Where:

- d_j is the Jaro distance defined above
- L is the length of the common prefix up to the first four characters

For example, consider the strings

s_1 = FRANKLIN

s_2 = FRAKNLIN

When looking for common characters, they must exist in s_2 no further than 8/2-1 = 3 spaces away from their position in s_1. Thus, c = 8. Further, there is a transposition of two character pairs, so t = 2/2 = 1.

The Jaro distance d_j = 1/3(8/8 + 8/8 + (8-1)/8) = 0.9583.

Next, we find L, the length of the common prefix. In this example, L = 3 (“FRA”).

Therefore, the Jaro-Winkler distance d_{jw} = 0.9583 + 3/10(1-0.9583) = 0.97081.

In its current implementation, the Jaro-Winkler algorithm is further modified to allow for common transpositions (see list below). These will be treated as matches as we search for common characters. For example, if we were comparing “JONES” and “J0NES” (a zero in place of an “O”), the number of common characters would be 5.

Appendix Table-1 List of Common Transpositions Used in Jaro-Winkler Algorithm

From	To
A	E
A	I
A	O
A	U
B	V
E	I
E	O
E	U
I	O
I	U
O	U
I	Y
E	Y
C	G
E	F
W	U
W	V
X	K
S	Z
X	S
Q	C
U	V
M	N
L	I
Q	O
P	R
I	J
2	Z
5	S
8	B
1	I
1	L
0	O
C	K
G	J
E	
Y	
S	

Another option that is available further increases the distance measure d_{jw} if the string is “long”. This option has been turned off for purposes of the dyad linking system.

The code used for the dyad linking system is almost identical to the code as originally written by Winkler[11] with very minor modifications to allow for its compilation in C++.

[11] Original C Implementation: http://web.archive.org/web/20100227020019/http://www.census.gov/geo/msb/stand/strcmp.c

Since there are many cases where name fields do not contain commas, our chances of inaccurately assigning last name are fairly high. Additionally, since the Jaro-Winkler distance is particularly sensitive to the beginning of the string, several permutations of first, middle and last name fields are considered. To reduce errors, when constructing name strings no spaces are used. If we do not find a suitable match to an observation using the order last name, first name and middle name then the following permutations are considered (until a suitable match is found):

- first middle last
- middle first last
- last first
- first last

Additionally, there are considerations taken when deciding if we should use the middle name field when constructing our strings for comparison. Let us examine a simple example. If in the first dataset, name = SMITHJOHNJACOB and in the second, name = SMITHJOHN and we were to use the full name in the first dataset and compare it to the full name in the second, our Jaro-Winkler distance would be as follows:

$d_j = 1/3(9/14 + 9/9 + 1) = 0.881$

$d_{jw} = 0.881 + 4/10(1-0.881) = 0.9286$

The mere fact that the first string is much longer than the second adversely affects the score. If however, we removed middle name from consideration in the first string, our value of d_{jw} would be 1.

It is not enough to simple check for the presence of the middle name in both strings. We also need to account for cases where in the first record the name has three parts JOHNJACOBSMITH and in the second name, only two parts JOHNJACOBSMITH, where the last name was cleaned from a hyphenated last name. These sorts of situations are common when dealing with Hispanic and Native American names. Data entry for such names is frequently inconsistent.

Thus, given two persons we wish to compare and create name1 for the first person and name2 for the second, strings are created as follows:

A middle name is used in name2 if:

- person 1 has a middle name OR
- length of last name for person 1 > length of last name for person 2 AND person 1 has a middle name

Similarly, a middle name is used in name1 if:

- person 2 has a middle name OR
- length of last name for person 2 > length of last name for person 1 AND person 2 has a middle name

In the simple case where middle name exists in both, this results in middle name being used in both name1 and name2. If middle name only exists in one person-record, and not in the other and last names are identical then middle name is disregarded. However, if last names have different lengths, middle name is used.

B. Detailed Results

1. EOUSA Matters OUT/USMS IN

Appendix Table-2 Detailed Results EOUSA Matters OUT/USMS IN: Overall Link Rate

	% EOUSA linked	% USMS linked
1994	47.26%	69.95%
1995	50.78%	74.71%
1996	53.25%	73.76%
1997	56.47%	74.28%
1998	60.26%	73.71%
1999	60.31%	75.11%
2000	59.86%	73.39%
2001	59.91%	73.69%
2002	60.38%	75.41%
2003	60.27%	76.21%
2004	62.59%	77.21%
2005	63.18%	77.77%
2006	64.83%	74.41%
2007	66.88%	73.77%
2008	73.69%	89.13%
2009	72.41%	88.72%

Appendix Table-3 Detailed Results EOUSA Matters OUT/USMS IN: Links by USMS Year

	eoYear																
msYear	**0**	**1994**	**1995**	**1996**	**1997**	**1998**	**1999**	**2000**	**2001**	**2002**	**2003**	**2004**	**2005**	**2006**	**2007**	**2008**	**2009**
0	0	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
1994	20,170	42,374	3,466	593	255	85	65	31	38	20	7	4	4	5	4	2	2
1995	17,368	2,656	44,143	3,492	550	154	92	61	62	26	49	4	10	5	1	8	2
1996	18,441	272	2,813	44,420	3,428	421	240	98	51	37	24	8	10	4	3	3	1
1997	19,440	130	265	2,780	47,247	4,269	526	198	117	74	490	20	11	4	4	1	7
1998	22,494	65	113	267	2,962	54,053	4,590	506	200	150	82	46	12	9	3	9	3
1999	22,336	31	61	80	260	3,290	58,139	4,388	496	289	220	64	36	17	6	5	3
2000	25,079	34	46	65	92	251	3,525	59,444	4,494	605	341	137	64	32	23	12	9
2001	25,319	22	36	38	60	101	305	3,790	60,673	4,817	629	226	102	65	26	20	10
2002	24,193	13	13	22	33	45	157	342	3,754	63,513	5,230	628	210	125	55	26	20
2003	23,843	18	17	15	34	29	64	111	273	3,938	65,700	5,183	542	238	126	75	41
2004	25,508	12	14	9	20	26	33	64	115	307	3,397	78,966	2,710	450	188	103	49
2005	24,423	5	6	13	10	23	31	41	71	121	318	4,515	77,247	2,329	451	160	108
2006	29,895	3	7	8	8	13	17	21	44	69	120	429	4,595	78,581	2,454	412	189
2007	32,618	2	7	6	6	17	11	24	38	36	63	160	354	4,599	83,298	2,749	399
2008	15,866	6	9	6	6	18	21	16	24	37	65	85	152	402	4,587	121,550	3,064
2009	17,491	3	2	6	3	4	8	14	16	25	34	53	85	154	392	4,384	132,441

Appendix Table-4 Detailed Results EOUSA Matters OUT/USMS IN: Links by EOUSA Year

	msYear																
eoYear	**0**	**1994**	**1995**	**1996**	**1997**	**1998**	**1999**	**2000**	**2001**	**2002**	**2003**	**2004**	**2005**	**2006**	**2007**	**2008**	**2009**
0	0	20,170	17,368	18,441	19,440	22,494	22,336	25,079	25,319	24,193	23,843	25,508	24,423	29,895	32,618	15,866	17,491
1994	52,199	43,480	2,679	278	132	65	32	34	22	13	19	12	5	3	2	6	3
1995	51,040	3,745	45,419	2,886	272	116	65	47	38	14	17	14	6	9	7	9	2
1996	46,735	680	3,769	45,442	2,808	275	81	66	40	22	15	9	13	8	6	6	6
1997	43,956	301	669	4,120	48,402	2,991	268	95	61	35	35	21	10	8	6	6	3
1998	42,311	99	190	486	4,483	55,035	3,317	256	102	47	30	26	23	13	17	18	4
1999	45,765	86	123	281	679	4,908	59,241	3,555	310	158	67	33	34	17	12	23	8
2000	47,629	56	83	126	264	570	4,785	60,673	3,831	356	112	70	41	22	24	16	14
2001	48,207	50	79	70	148	242	553	4,731	61,766	3,791	278	117	74	47	40	25	16
2002	49,708	32	35	48	90	172	332	684	5,167	64,606	3,962	313	122	69	39	37	28
2003	51,702	23	58	36	525	104	268	375	706	5,520	66,773	3,424	323	122	66	66	38
2004	56,037	4	6	11	28	55	85	168	258	735	5,580	81,549	4,577	439	162	87	54
2005	53,407	5	13	16	16	18	49	76	131	238	622	5,340	79,863	4,682	366	155	88
2006	50,142	5	5	5	7	17	20	39	73	140	273	676	4,840	81,091	4,696	410	158
2007	48,160	7	3	3	5	5	8	27	30	63	151	272	656	4,664	86,286	4,668	401
2008	48,481	2	8	3	1	13	6	15	22	34	85	143	217	595	4,941	125,213	4,459
2009	53,742	3	3	1	9	3	6	11	10	25	49	80	134	237	576	5,336	134,605

Appendix Table-5 Detailed Results EOUSA Matters OUT/USMS IN: Links by Block

	Total Linked	**Linked by Block**	**% of Total Linked by Block**
block 1	904,192	904,192	69.95%
block 2	1,132,353	228,161	17.65%
block 3	1,285,167	152,814	11.82%
block 4	1,292,667	7,500	0.58%

2. AOUSC IN/EOUSA IN

Appendix Table-6 Detailed Results AOUSC IN/EOUSA Cases IN: Overall Link Rate

	% AOUSC linked	% EOUSA linked
1994	74.25%	76.51%
1995	79.27%	78.23%
1996	78.51%	79.26%
1997	75.16%	75.47%
1998	76.97%	87.38%
1999	79.24%	87.67%
2000	80.95%	86.90%
2001	81.03%	87.07%
2002	80.40%	86.49%
2003	78.87%	87.98%
2004	80.10%	80.99%
2005	79.04%	76.86%
2006	80.22%	76.65%
2007	80.31%	79.33%
2008	82.64%	80.56%
2009	81.60%	79.98%

Appendix Table-7 Detailed Results AOUSC IN/EOUSA Cases IN: Links by EOUSA Year

									eoYear								
aoYear	**0**	**1994**	**1995**	**1996**	**1997**	**1998**	**1999**	**2000**	**2001**	**2002**	**2003**	**2004**	**2005**	**2006**	**2007**	**2008**	**2009**
0	0	14,272	14,127	13,786	17,710	8,771	9,014	10,212	10,032	10,965	9,849	17,660	21,946	21,520	18,695	18,375	19,837
1994	16,277	46,031	790	68	32	5	6	0	0	0	0	0	0	0	0	0	0
1995	13,332	373	49,536	1,027	69	16	1	0	0	0	0	0	0	0	0	0	0
1996	14,247	1	290	50,869	848	41	5	0	0	0	0	0	0	0	0	0	0
1997	17,407	36	97	566	51,531	470	7	0	0	0	0	0	0	0	0	0	0
1998	18,167	18	51	98	1,785	58,559	215	0	0	0	0	0	0	0	0	0	0
1999	16,753	6	16	26	153	1,473	62,154	156	0	0	0	0	0	0	0	0	0
2000	15,975	7	11	10	50	104	1,594	65,781	318	10	4	21	17	3	1	4	1
2001	15,782	4	3	4	21	42	106	1,645	65,289	244	13	19	16	8	1	2	2
2002	17,306	4	3	7	14	17	67	147	1,830	68,541	307	28	7	12	2	4	1
2003	19,573	2	1	2	4	2	3	20	109	1,321	71,094	489	16	17	5	3	4
2004	18,565	2	2	1	0	3	5	13	14	74	645	73,610	332	17	7	3	3
2005	19,321	0	0	0	0	0	0	8	8	31	53	1,004	71,444	279	13	6	5
2006	17,427	0	0	0	0	0	5	5	6	16	18	36	1,025	69,309	283	20	17
2007	17,573	0	0	0	0	2	0	3	3	0	8	15	22	967	70,384	272	19
2008	16,028	0	2	1	0	2	1	0	2	2	1	1	12	43	1,012	74,961	255
2009	18,019	1	0	0	0	1	1	0	1	1	2	6	9	18	34	892	78,980

Appendix Table-8 Detailed Results AOUSC IN/EOUSA Cases IN: Links by AOUSC Year

									aoYear							
eoYear	**0**	**1994**	**1995**	**1996**	**1997**	**1998**	**1999**	**2000**	**2001**	**2002**	**2003**	**2004**	**2005**	**2006**	**2007**	**2008**	**2009**
0	0	16,277	13,332	14,247	17,407	18,167	16,753	15,975	15,782	17,306	19,573	18,565	19,321	17,427	17,573	16,028	18,019
1994	14,272	46,031	373	1	36	18	6	7	4	4	2	2	0	0	0	0	1
1995	14,127	790	49,536	290	97	51	16	11	3	3	1	2	0	0	0	2	0
1996	13,786	68	1,027	50,869	566	98	26	10	4	7	2	1	0	0	0	1	0
1997	17,710	32	69	848	51,531	1,785	153	50	21	14	4	0	0	0	0	0	0
1998	8,771	5	16	41	470	58,559	1,473	104	42	17	2	3	0	0	2	2	1
1999	9,014	6	1	5	7	215	62,154	1,594	106	67	3	5	0	5	0	1	1
2000	10,212	0	0	0	0	0	156	65,781	1,645	147	20	13	8	5	3	0	0
2001	10,032	0	0	0	0	0	0	318	65,289	1,830	109	14	8	6	3	2	1
2002	10,965	0	0	0	0	0	0	10	244	68,541	1,321	74	31	16	0	2	1
2003	9,849	0	0	0	0	0	0	4	13	307	71,094	645	53	18	8	1	2
2004	17,660	0	0	0	0	0	0	21	19	28	489	73,610	1,004	36	15	1	6
2005	21,946	0	0	0	0	0	0	17	16	7	16	332	71,444	1,025	22	12	9
2006	21,520	0	0	0	0	0	0	3	8	12	17	17	279	69,309	967	43	18
2007	18,695	0	0	0	0	0	0	1	1	2	5	7	13	283	70,384	1,012	34
2008	18,375	0	0	0	0	0	0	4	2	4	3	3	6	20	272	74,961	892
2009	19,837	0	0	0	0	0	0	1	2	1	4	3	5	17	19	255	78,980

Appendix Table-9 Detailed Results AOUSC IN/EOUSA Cases IN: Links by Block

	Total Linked	Linked by Block	% of Total Linked by Block
block 1	901,802	901,802	85.63%
block 2	1,021,425	119,623	11.36%
block 3	1,049,230	27,805	2.64%
block 4	1,053,121	3,891	0.37%

3. AOUSC OUT/EOUSA OUT

Appendix Table-10 Detailed Results AOUSC OUT/EOUSA Cases OUT and Magistrate Matters OUT: Overall Link Rate

	% AOUSC linked	% EOUSA linked
1994	75.07%	68.74%
1995	78.95%	66.88%
1996	80.26%	72.74%
1997	80.60%	70.76%
1998	78.92%	68.72%
1999	80.21%	68.93%
2000	83.46%	68.93%
2001	83.55%	68.25%
2002	85.79%	70.39%
2003	84.29%	70.23%
2004	84.24%	62.99%
2005	83.19%	63.70%
2006	84.36%	62.55%
2007	82.76%	61.86%
2008	84.92%	49.26%
2009	85.92%	49.47%

Appendix Table-11 Detailed Results AOUSC OUT/EOUSA Cases OUT and Magistrate Matters OUT: Links by EOUSA Year

									eoYea								
aoYear	**0**	**1994**	**1995**	**1996**	**1997**	**1998**	**1999**	**2000**	**2001**	**2002**	**2003**	**2004**	**2005**	**2006**	**2007**	**2008**	**2009**
0	0	21,504	22,731	18,876	21,887	25,316	27,807	29,032	30,183	29,265	30,731	41,929	41,519	44,795	45,386	80,657	84,288
1994	15,859	46,371	1,096	191	66	25	16	0	0	0	0	0	0	0	0	0	0
1995	12,154	787	43,700	828	193	42	22	0	0	0	0	0	0	0	0	0	0
1996	12,424	96	1,006	48,365	897	123	40	0	0	0	0	0	0	0	0	0	0
1997	12,771	26	74	895	51,100	837	114	0	0	0	0	0	0	0	0	0	0
1998	14,867	3	18	55	652	54,040	888	0	0	0	0	0	0	0	0	0	0
1999	15,140	9	14	30	58	548	60,185	508	0	0	0	0	0	0	0	0	0
2000	12,843	0	0	1	1	1	437	63,353	828	96	31	20	17	11	4	3	4
2001	12,803	0	1	0	0	0	0	479	63,586	737	96	37	30	27	4	10	5
2002	11,528	0	0	0	0	0	0	28	429	68,172	757	118	46	21	15	7	7
2003	13,501	0	0	0	0	0	1	10	22	497	71,036	678	91	52	25	12	7
2004	13,255	0	0	0	0	0	0	6	10	28	521	69,492	602	130	35	18	13
2005	14,671	0	0	0	0	0	0	7	8	9	25	863	70,914	656	71	39	31
2006	13,877	0	0	0	0	0	0	3	4	8	18	98	1,061	72,845	690	68	35
2007	15,339	0	0	0	1	1	0	4	1	3	13	28	56	964	71,808	667	76
2008	13,897	0	0	0	0	0	0	1	3	2	6	20	20	76	880	76,596	648
2009	13,553	0	0	0	0	0	0	3	0	4	4	6	9	41	76	880	81,701

Appendix Table-12 Detailed Results AOUSC OUT/EOUSA Cases OUT and Magistrate Matters OUT: Links by AOUSC Year

									aoYear							
eoYear	**0**	**1994**	**1995**	**1996**	**1997**	**1998**	**1999**	**2000**	**2001**	**2002**	**2003**	**2004**	**2005**	**2006**	**2007**	**2008**	**2009**
0	0	15,859	12,154	12,424	12,771	14,867	15,140	12,843	12,803	11,528	13,501	13,255	14,671	13,877	15,339	13,897	13,553
1994	21,504	46,371	787	96	26	3	9	0	0	0	0	0	0	0	0	0	0
1995	22,731	1,096	43,700	1,006	74	18	14	0	1	0	0	0	0	0	0	0	0
1996	18,876	191	828	48,365	895	55	30	1	0	0	0	0	0	0	0	0	0
1997	21,887	66	193	897	51,100	652	58	1	0	0	0	0	0	0	1	0	0
1998	25,316	25	42	123	837	54,040	548	1	0	0	0	0	0	0	1	0	0
1999	27,807	16	22	40	114	888	60,185	437	0	0	1	0	0	0	0	0	0
2000	29,032	0	0	0	0	0	508	63,353	479	28	10	6	7	3	4	1	3
2001	30,183	0	0	0	0	0	0	828	63,586	429	22	10	8	4	1	3	0
2002	29,265	0	0	0	0	0	0	96	737	68,172	497	28	9	8	3	2	4
2003	30,731	0	0	0	0	0	0	31	96	757	71,036	521	25	18	13	6	4
2004	41,929	0	0	0	0	0	0	20	37	118	678	69,492	863	98	28	20	6
2005	41,519	0	0	0	0	0	0	17	30	46	91	602	70,914	1,061	56	20	9
2006	44,795	0	0	0	0	0	0	11	27	21	52	130	656	72,845	964	76	41
2007	45,386	0	0	0	0	0	0	4	4	15	25	35	71	690	71,808	880	76
2008	80,657	0	0	0	0	0	0	3	10	7	12	18	39	68	667	76,596	880
2009	84,288	0	0	0	0	0	0	4	5	7	7	13	31	35	76	648	81,701

Appendix Table-13 Detailed Results AOUSC OUT/EOUSA Cases OUT and Magistrate Matters OUT: Links by Block

	Total Linked	Linked by Block	% of Total Linked by Block
block 1	792,509	792,509	76.30%
block 2	978,215	185,706	17.88%
block 3	1,033,328	55,113	5.31%
block 4	1,038,674	5,346	0.51%

4. AOUSC OUT/USSC OUT

Appendix Table-14 Detailed Results AOUSC OUT/USSC OUT: Overall Link Rate

	% AOUSC linked	%USSC linked
1994	72.95%	58.13%
1995	72.67%	59.87%
1996	74.69%	62.97%
1997	80.10%	68.85%
1998	75.83%	65.55%
1999	77.08%	66.56%
2000	80.33%	70.51%
2001	80.26%	70.69%
2002	82.82%	73.30%
2003	85.48%	75.46%
2004	85.91%	76.38%
2005	84.97%	75.97%
2006	82.99%	74.75%
2007	85.38%	76.16%
2008	86.44%	77.69%
2009	87.51%	79.05%

Appendix Table-15 Detailed Results AOUSC OUT/USSC OUT: Links by USSC Year

	scYear																
aoYear	**0**	**1994**	**1995**	**1996**	**1997**	**1998**	**1999**	**2000**	**2001**	**2002**	**2003**	**2004**	**2005**	**2006**	**2007**	**2008**	**2009**
0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
1994	13,715	36,986	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
1995	12,998	0	34,558	0	0	0	0	0	0	0	0	0	0	0	0	0	0
1996	13,433	0	0	39,643	0	0	0	0	0	0	0	0	0	0	0	0	0
1997	11,257	0	0	0	45,313	0	0	0	0	0	0	0	0	0	0	0	0
1998	14,733	0	0	0	0	46,225	0	0	0	0	0	0	0	0	0	0	0
1999	15,139	0	0	0	0	0	50,916	0	0	0	0	0	0	0	0	0	0
2000	13,408	0	0	0	0	0	0	54,748	0	0	0	0	0	0	0	0	0
2001	13,527	0	0	0	0	0	0	0	55,006	0	0	0	0	0	0	0	0
2002	12,335	0	0	0	0	0	0	0	0	59,463	0	0	0	0	0	0	0
2003	11,017	0	0	0	0	0	0	0	0	0	64,842	0	0	0	0	0	0
2004	10,539	0	0	0	0	0	0	0	0	0	0	64,243	0	0	0	0	0
2005	11,726	0	0	0	0	0	0	0	0	0	0	0	66,316	0	0	0	0
2006	13,593	0	0	0	0	0	0	0	0	0	0	0	0	66,311	0	0	0
2007	11,605	0	0	0	0	0	0	0	0	0	0	0	0	0	67,751	0	0
2008	11,228	0	0	0	0	0	0	0	0	0	0	0	0	0	0	71,595	0
2009	10,865	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	76,110

Appendix Table-16 Detailed Results AOUSC OUT/USSC OUT: Links by AOUSC Year

	aoYear																
scYear	**0**	**1994**	**1995**	**1996**	**1997**	**1998**	**1999**	**2000**	**2001**	**2002**	**2003**	**2004**	**2005**	**2006**	**2007**	**2008**	**2009**
0	0	26,638	23,168	23,308	20,504	24,298	25,576	22,902	22,809	21,665	21,090	19,867	20,978	22,396	21,210	20,554	20,167
1994	3,552	36,986	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
1995	3,965	0	34,558	0	0	0	0	0	0	0	0	0	0	0	0	0	0
1996	3,390	0	0	39,643	0	0	0	0	0	0	0	0	0	0	0	0	0
1997	3,535	0	0	0	45,313	0	0	0	0	0	0	0	0	0	0	0	0
1998	4,529	0	0	0	0	46,225	0	0	0	0	0	0	0	0	0	0	0
1999	4,641	0	0	0	0	0	50,916	0	0	0	0	0	0	0	0	0	0
2000	5,098	0	0	0	0	0	0	54,748	0	0	0	0	0	0	0	0	0
2001	4,891	0	0	0	0	0	0	0	55,006	0	0	0	0	0	0	0	0
2002	4,903	0	0	0	0	0	0	0	0	59,463	0	0	0	0	0	0	0
2003	5,416	0	0	0	0	0	0	0	0	0	64,842	0	0	0	0	0	0
2004	5,825	0	0	0	0	0	0	0	0	0	0	64,243	0	0	0	0	0
2005	6,146	0	0	0	0	0	0	0	0	0	0	0	66,316	0	0	0	0
2006	6,274	0	0	0	0	0	0	0	0	0	0	0	0	66,311	0	0	0
2007	5,114	0	0	0	0	0	0	0	0	0	0	0	0	0	67,751	0	0
2008	4,883	0	0	0	0	0	0	0	0	0	0	0	0	0	0	71,595	0
2009	5,262	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	76,110

Appendix Table-17 Detailed Results AOUSC OUT/USSC OUT: Links by Block

	Total Linked	Linked by Block	% of Total Linked by Block
block 1	662,747	662,747	73.64%
block 2	898,731	235,984	26.22%
block 3	900,026	1,295	0.14%

5. BOP IN/USSC OUT

Appendix Table-18 Detailed Results BOP IN/USSC OUT: Overall Link Rate

	% BOP linked	% USSC linked
1994	67.01%	88.13%
1995	78.75%	88.01%
1996	81.32%	87.22%
1997	84.83%	86.08%
1998	82.50%	85.84%
1999	80.72%	84.35%
2000	82.49%	84.25%
2001	82.07%	85.27%
2002	86.62%	83.72%
2003	89.51%	82.86%
2004	87.64%	82.59%
2005	87.09%	82.60%
2006	84.41%	82.71%
2007	85.04%	81.37%
2008	86.57%	78.23%
2009	86.86%	59.85%

Appendix Table-19 Detailed Results BOP IN/USSC OUT: Links by USSC Year

	scYear																
bopYear	**0**	**1994**	**1995**	**1996**	**1997**	**1998**	**1999**	**2000**	**2001**	**2002**	**2003**	**2004**	**2005**	**2006**	**2007**	**2008**	**2009**
0	0	12,857	11,941	12,759	15,421	14,742	16,849	17,690	17,052	18,797	20,890	20,458	20,227	19,934	21,015	23,786	38,186
1994	11,367	23,089	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
1995	6,929	3,750	21,926	0	0	0	0	0	0	0	0	0	0	0	0	0	0
1996	6,595	339	3,894	24,473	0	0	0	0	0	0	0	0	0	0	0	0	0
1997	5,835	156	298	4,977	27,191	0	0	0	0	0	0	0	0	0	0	0	0
1998	7,477	100	144	379	5,390	29,243	0	0	0	0	0	0	0	0	0	0	0
1999	9,184	67	98	153	367	5,988	31,776	0	0	0	0	0	0	0	0	0	0
2000	8,753	39	50	84	177	315	6,085	34,493	0	0	0	0	0	0	0	0	0
2001	9,217	32	45	69	91	147	380	6,842	34,588	0	0	0	0	0	0	0	0
2002	6,797	24	41	37	71	102	172	356	7,315	35,873	0	0	0	0	0	0	0
2003	5,655	21	15	27	50	70	99	191	460	8,822	38,513	0	0	0	0	0	0
2004	7,016	15	25	20	22	55	68	99	199	439	9,919	38,871	0	0	0	0	0
2005	7,683	20	14	17	24	22	46	72	118	197	505	9,873	40,928	0	0	0	0
2006	9,605	11	9	14	21	24	38	45	68	109	207	501	10,449	40,514	0	0	0
2007	9,196	7	8	10	10	19	16	21	44	57	105	196	493	11,318	39,958	0	0
2008	8,300	8	9	8	11	17	14	21	31	42	71	115	220	564	11,324	41,049	0
2009	8,473	3	6	6	2	10	14	16	22	30	48	54	145	255	568	11,643	43,186

Appendix Table-20 Detailed Results BOP IN/USSC OUT: Links by BOP Year

									bopYear								
scYear	**0**	**1994**	**1995**	**1996**	**1997**	**1998**	**1999**	**2000**	**2001**	**2002**	**2003**	**2004**	**2005**	**2006**	**2007**	**2008**	**2009**
1994	3,729	23,089	3,750	339	156	100	67	39	32	24	21	15	20	11	7	8	3
1995	3,622	0	21,926	3,894	298	144	98	50	45	41	15	25	14	9	8	9	6
1996	4,437	0	0	24,473	4,977	379	153	84	69	37	27	20	17	14	10	8	6
1997	5,404	0	0	0	27,191	5,390	367	177	91	71	50	22	24	21	10	11	2
1998	5,939	0	0	0	0	29,243	5,988	315	147	102	70	55	22	24	19	17	10
1999	7,180	0	0	0	0	0	31,776	6,085	380	172	99	68	46	38	16	14	14
2000	7,878	0	0	0	0	0	0	34,493	6,842	356	191	99	72	45	21	21	16
2001	7,404	0	0	0	0	0	0	0	34,588	7,315	460	199	118	68	44	31	22
2002	8,859	0	0	0	0	0	0	0	0	35,873	8,822	439	197	109	57	42	30
2003	10,213	0	0	0	0	0	0	0	0	0	38,513	9,919	505	207	105	71	48
2004	10,458	0	0	0	0	0	0	0	0	0	0	38,871	9,873	501	196	115	54
2005	11,007	0	0	0	0	0	0	0	0	0	0	0	40,928	10,449	493	220	145
2006	11,005	0	0	0	0	0	0	0	0	0	0	0	0	40,514	11,318	564	255
2007	11,875	0	0	0	0	0	0	0	0	0	0	0	0	0	39,958	11,324	568
2008	14,662	0	0	0	0	0	0	0	0	0	0	0	0	0	0	41,049	11,643
2009	28,976	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	43,186

Appendix Table-21 Detailed Results BOP IN/USSC OUT: Links by Block

	Total Linked	Linked by Block	% of Total Linked by Block
block 1	604,819	604,819	89.62%
block 2	663,005	58,186	8.62%
block 3	671,596	8,591	1.27%
block 4	674,846	3,250	0.48%

6. AOUSC IN/AOUSC OUT

Appendix Table-22 Detailed Results AOUSC IN/AOUSC OUT: Overall Link Rate

	% AOUSC Cases In	% AOUSC Cases Out
1994	92.37%	48.31%
1995	94.31%	86.10%
1996	92.14%	93.53%
1997	91.94%	95.05%
1998	93.34%	95.59%
1999	93.28%	95.52%
2000	92.70%	96.12%
2001	93.12%	96.78%
2002	91.64%	96.59%
2003	90.01%	95.45%
2004	89.36%	94.33%
2005	87.61%	92.48%
2006	87.57%	90.44%
2007	86.18%	90.41%
2008	79.86%	90.87%
2009	41.05%	91.61%

Appendix Table-23 Detailed Results AOUSC IN/AOUSC OUT: Links by AOUSC OUT Year

	aoOutYear																
aoInYear	**0**	**1994**	**1995**	**1996**	**1997**	**1998**	**1999**	**2000**	**2001**	**2002**	**2003**	**2004**	**2005**	**2006**	**2007**	**2008**	**2009**
0	0	32,886	8,025	4,070	3,258	3,112	3,424	3,013	2,506	2,766	3,913	4,765	6,563	8,480	8,530	8,414	8,075
1994	4,826	30,690	21,558	3,485	933	400	540	202	103	114	114	44	38	33	57	46	26
1995	3,662	7	28,057	25,916	4,052	1,167	603	258	144	128	93	38	49	46	43	52	39
1996	5,213	1	2	29,379	25,371	3,603	1,197	500	231	186	269	81	62	54	67	54	31
1997	5,648	17	48	52	32,103	25,685	4,012	1,094	398	241	341	138	102	75	57	64	39
1998	5,255	14	16	31	42	36,425	29,983	4,398	1,097	542	376	208	111	108	95	113	79
1999	5,425	3	6	9	36	72	36,639	30,904	4,591	1,496	652	278	174	139	132	96	85
2000	6,123	3	3	6	12	39	41	37,191	32,193	5,261	1,509	568	278	209	215	133	127
2001	5,723	1	3	1	3	8	34	29	36,458	33,129	4,967	1,294	621	337	320	139	134
2002	7,383	0	2	0	3	9	10	34	48	37,189	35,095	5,212	1,587	754	575	213	183
2003	9,261	1	1	1	0	0	1	20	27	42	38,516	34,973	6,263	1,934	978	374	273
2004	9,927	0	0	0	2	1	1	1	9	18	28	36,385	36,826	6,756	2,045	870	427
2005	11,423	0	0	0	0	2	4	5	6	4	47	68	34,518	36,342	6,899	1,958	896
2006	10,957	0	1	0	0	0	1	1	2	3	10	38	45	33,339	35,274	6,481	2,015
2007	12,341	0	1	1	1	0	0	0	1	5	0	10	35	44	33,579	36,613	6,637
2008	18,595	1	1	0	1	0	1	0	1	2	2	6	14	43	57	36,488	37,111
2009	57,755	0	2	0	0	0	1	0	0	2	0	4	8	14	38	41	40,100

Appendix Table-24 Detailed Results AOUSC IN/AOUSC OUT: Links by AOUSC IN Year

									aoInYear								
aoOutYear	**0**	**1994**	**1995**	**1996**	**1997**	**1998**	**1999**	**2000**	**2001**	**2002**	**2003**	**2004**	**2005**	**2006**	**2007**	**2008**	**2009**
0	0	4,826	3,662	5,213	5,648	5,255	5,425	6,123	5,723	7,383	9,261	9,927	11,423	10,957	12,341	18,595	57,755
1994	32,886	30,690	7	1	17	14	3	3	1	0	1	0	0	0	0	1	0
1995	8,025	21,558	28,057	2	48	16	6	3	3	2	1	0	0	1	1	1	2
1996	4,070	3,485	25,916	29,379	52	31	9	6	1	0	1	0	0	0	1	0	0
1997	3,258	933	4,052	25,371	32,103	42	36	12	3	3	0	2	0	0	1	1	0
1998	3,112	400	1,167	3,603	25,685	36,425	72	39	8	9	0	1	2	0	0	0	0
1999	3,424	540	603	1,197	4,012	29,983	36,639	41	34	10	1	1	4	1	0	1	1
2000	3,013	202	258	500	1,094	4,398	30,904	37,191	29	34	20	1	5	1	0	0	0
2001	2,506	103	144	231	398	1,097	4,591	32,193	36,458	48	27	9	6	2	1	1	0
2002	2,766	114	128	186	241	542	1,496	5,261	33,129	37,189	42	18	4	3	5	2	2
2003	3,913	114	93	269	341	376	652	1,509	4,967	35,095	38,516	28	47	10	0	2	0
2004	4,765	44	38	81	138	208	278	568	1,294	5,212	34,973	36,385	68	38	10	6	4
2005	6,563	38	49	62	102	111	174	278	621	1,587	6,263	36,826	34,518	45	35	14	8
2006	8,480	33	46	54	75	108	139	209	337	754	1,934	6,756	36,342	33,339	44	43	14
2007	8,530	57	43	67	57	95	132	215	320	575	978	2,045	6,899	35,274	33,579	57	38
2008	8,414	46	52	54	64	113	96	133	139	213	374	870	1,958	6,481	36,613	36,488	41
2009	8,075	26	39	31	39	79	85	127	134	183	273	427	896	2,015	6,637	37,111	40,100

7. EOUSA Cases IN/EOUSA Cases OUT

Appendix Table-25 Detailed Results EOUSA Cases IN/EOUSA Cases OUT: Overall Link Rate

	% EOUSA Cases In	% EOUSA Cases Out
1994	92.20%	38.81%
1995	89.83%	86.11%
1996	86.72%	94.69%
1997	86.74%	89.95%
1998	92.17%	89.03%
1999	93.50%	86.34%
2000	92.71%	86.76%
2001	92.82%	86.48%
2002	92.74%	85.75%
2003	89.27%	86.57%
2004	91.16%	89.67%
2005	90.56%	94.35%
2006	89.73%	94.48%
2007	87.36%	94.43%
2008	79.39%	95.29%
2009	40.66%	95.98%

Appendix Table-26 Detailed Results EOUSA Cases IN/EOUSA Cases OUT: Links by EOUSA Cases OUT Year

									eoCOutYear								
eoCInYear	**0**	**1994**	**1995**	**1996**	**1997**	**1998**	**1999**	**2000**	**2001**	**2002**	**2003**	**2004**	**2005**	**2006**	**2007**	**2008**	**2009**
0	0	36,125	8,048	3,218	6,514	7,534	10,243	10,529	10,947	12,070	11,873	8,618	4,872	5,039	4,899	4,350	3,827
1994	4,737	22,913	26,496	4,388	968	364	215	142	97	118	85	68	32	29	43	26	36
1995	6,603	2	23,385	28,226	4,454	1,065	371	245	147	93	87	45	35	59	39	44	29
1996	8,829	2	1	24,725	26,962	3,587	1,046	422	243	189	154	65	50	55	43	51	41
1997	9,577	0	0	0	25,633	28,913	5,360	1,202	433	283	180	153	131	128	93	78	53
1998	5,445	0	0	0	2	26,994	29,127	5,055	1,156	556	281	207	152	163	122	126	122
1999	4,756	0	0	0	1	5	28,341	31,480	5,309	1,572	568	301	226	210	158	164	93
2000	5,684	0	0	0	1	3	2	30,123	32,971	5,823	1,489	705	377	272	199	189	152
2001	5,571	0	0	0	0	0	0	2	29,270	33,775	5,672	1,461	738	446	299	196	182
2002	5,896	0	0	0	0	0	0	2	1	29,744	36,160	5,343	1,915	1,072	548	317	207
2003	8,800	0	0	0	0	0	0	0	1	1	30,845	30,844	6,790	2,583	1,196	551	383
2004	8,208	0	0	0	53	56	58	75	100	122	356	33,770	35,928	9,282	2,965	1,245	671
2005	8,956	0	0	0	59	50	60	77	104	114	174	811	33,520	37,610	8,914	3,031	1,366
2006	9,472	0	0	0	76	44	53	64	77	104	228	510	967	33,303	35,707	8,634	2,954
2007	11,428	0	0	0	45	22	24	38	45	69	114	211	248	593	31,818	37,230	8,552
2008	19,484	0	0	0	39	28	32	30	48	53	93	169	177	265	621	35,613	37,890
2009	58,823	0	0	0	41	25	33	32	45	42	69	119	140	204	304	554	38,693

Appendix Table-27 Detailed Results EOUSA Cases IN/EOUSA Cases OUT: Links by EOUSA Cases IN Year

									eoCInYear								
eoCOutYear	**0**	**1994**	**1995**	**1996**	**1997**	**1998**	**1999**	**2000**	**2001**	**2002**	**2003**	**2004**	**2005**	**2006**	**2007**	**2008**	**2009**
0	0	4,737	6,603	8,829	9,577	5,445	4,756	5,684	5,571	5,896	8,800	8,208	8,956	9,472	11,428	19,484	58,823
1994	36,125	22,913	2	2	0	0	0	0	0	0	0	0	0	0	0	0	0
1995	8,048	26,496	23,385	1	0	0	0	0	0	0	0	0	0	0	0	0	0
1996	3,218	4,388	28,226	24,725	0	0	0	0	0	0	0	0	0	0	0	0	0
1997	6,514	968	4,454	26,962	25,633	2	1	1	0	0	0	53	59	76	45	39	41
1998	7,534	364	1,065	3,587	28,913	26,994	5	3	0	0	0	56	50	44	22	28	25
1999	10,243	215	371	1,046	5,360	29,127	28,341	2	0	0	0	58	60	53	24	32	33
2000	10,529	142	245	422	1,202	5,055	31,480	30,123	2	2	0	75	77	64	38	30	32
2001	10,947	97	147	243	433	1,156	5,309	32,971	29,270	1	1	100	104	77	45	48	45
2002	12,070	118	93	189	283	556	1,572	5,823	33,775	29,744	1	122	114	104	69	53	42
2003	11,873	85	87	154	180	281	568	1,489	5,672	36,160	30,845	356	174	228	114	93	69
2004	8,618	68	45	65	153	207	301	705	1,461	5,343	30,844	33,770	811	510	211	169	119
2005	4,872	32	35	50	131	152	226	377	738	1,915	6,790	35,928	33,520	967	248	177	140
2006	5,039	29	59	55	128	163	210	272	446	1,072	2,583	9,282	37,610	33,303	593	265	204
2007	4,899	43	39	43	93	122	158	199	299	548	1,196	2,965	8,914	35,707	31,818	621	304
2008	4,350	26	44	51	78	126	164	189	196	317	551	1,245	3,031	8,634	37,230	35,613	554
2009	3,827	36	29	41	53	122	93	152	182	207	383	671	1,366	2,954	8,552	37,890	38,693

Appendix Table-28 Detailed Results EOUSA Cases IN/EOUSA Cases OUT: Links by Block

	Total Linked	**Linked by Block**	**% of Total Linked by Block**
block 1	1,063,685	1,063,685	96.03%
block 2	1,107,065	43,380	3.92%
block 3	1,107,623	558	0.05%

8. EOUSA Matters OUT/EOUSA Cases IN

Appendix Table-29 Detailed Results EOUSA Matters OUT/EOUSA Cases IN: Overall Link Rate

	% EOUSA Matters Out	% EOUSA Cases In
1994	51.83%	84.97%
1995	54.02%	86.32%
1996	57.55%	86.18%
1997	60.02%	84.24%
1998	61.27%	93.79%
1999	59.83%	94.21%
2000	61.84%	94.01%
2001	60.61%	93.90%
2002	61.02%	94.25%
2003	61.05%	96.80%
2004	57.24%	90.49%
2005	57.29%	87.75%
2006	56.33%	87.46%
2007	55.67%	89.46%
2008	46.51%	90.24%
2009	45.20%	90.51%

Appendix Table-30 Detailed Results EOUSA Matters OUT/EOUSA Cases IN: Links by EOUSA Matters OUT Year

	eoMOutYear															
eoCInYear	**0**	**1994**	**1995**	**1996**	**1997**	**1998**	**1999**	**2000**	**2001**	**2002**	**2003**	**2004**	**2005**	**2006**	**2007**	**2008**	**2009**
0	0	47,684	47,682	42,443	40,374	41,229	46,315	45,283	47,355	48,897	50,679	64,068	61,965	62,277	64,454	98,547	106,763
1994	9,129	51,247	326	55	0	0	0	0	0	0	0	0	0	0	0	0	0
1995	8,885	29	55,677	338	0	0	0	0	0	0	0	0	0	0	0	0	0
1996	9,185	24	21	57,145	90	0	0	0	0	0	0	0	0	0	0	0	0
1997	11,380	0	0	0	60,490	202	92	27	11	5	6	2	1	0	0	0	1
1998	4,313	0	0	0	10	64,954	146	47	18	6	10	1	1	2	0	0	0
1999	4,239	0	0	0	7	34	68,599	238	38	13	7	5	0	1	0	2	1
2000	4,672	0	0	0	4	18	107	72,997	126	40	7	5	8	3	1	2	0
2001	4,736	0	0	0	1	8	14	37	72,537	203	34	15	10	9	3	3	2
2002	4,670	0	0	0	1	2	3	21	94	76,171	172	40	12	9	4	2	4
2003	2,622	0	0	0	0	1	2	3	18	56	78,966	249	46	17	7	5	2
2004	8,830	0	0	0	0	1	7	3	13	32	164	82,880	513	139	112	104	91
2005	11,623	0	0	0	1	3	5	5	10	7	56	2,058	80,211	484	164	124	95
2006	11,563	0	0	0	4	3	0	3	7	3	15	269	1,867	77,777	463	133	86
2007	9,531	0	0	0	1	2	0	0	5	1	2	123	259	1,544	78,305	535	129
2008	9,230	0	0	0	1	0	5	6	1	7	6	64	120	213	1,688	82,746	455
2009	9,411	0	0	0	3	0	5	2	1	3	5	56	72	122	208	2,035	87,201

Appendix Table-31 Detailed Results EOUSA Matters OUT/EOUSA Cases IN: Links by EOUSA Cases IN Year

	eoCInYear															
eoMOutYear	**0**	**1994**	**1995**	**1996**	**1997**	**1998**	**1999**	**2000**	**2001**	**2002**	**2003**	**2004**	**2005**	**2006**	**2007**	**2008**	**2009**
0	0	9,129	8,885	9,185	11,380	4,313	4,239	4,672	4,736	4,670	2,622	8,830	11,623	11,563	9,531	9,230	9,411
1994	47,684	51,247	29	24	0	0	0	0	0	0	0	0	0	0	0	0	0
1995	47,682	326	55,677	21	0	0	0	0	0	0	0	0	0	0	0	0	0
1996	42,443	55	338	57,145	0	0	0	0	0	0	0	0	0	0	0	0	0
1997	40,374	0	0	90	60,490	10	7	4	1	1	0	0	1	4	1	1	3
1998	41,229	0	0	0	202	64,954	34	18	8	2	1	1	3	3	2	0	0
1999	46,315	0	0	0	92	146	68,599	107	14	3	2	7	5	0	0	5	5
2000	45,283	0	0	0	27	47	238	72,997	37	21	3	3	5	3	0	6	2
2001	47,355	0	0	0	11	18	38	126	72,537	94	18	13	10	7	5	1	1
2002	48,897	0	0	0	5	6	13	40	203	76,171	56	32	7	3	1	7	3
2003	50,679	0	0	0	6	10	7	7	34	172	78,966	164	56	15	2	6	5
2004	64,068	0	0	0	2	1	5	5	15	40	249	82,880	2,058	269	123	64	56
2005	61,965	0	0	0	1	1	0	8	10	12	46	513	80,211	1,867	259	120	72
2006	62,277	0	0	0	0	2	1	3	9	9	17	139	484	77,777	1,544	213	122
2007	64,454	0	0	0	0	0	0	1	3	4	7	112	164	463	78,305	1,688	208
2008	98,547	0	0	0	0	0	2	2	3	2	5	104	124	133	535	82,746	2,035
2009	106,763	0	0	0	1	0	1	0	2	4	2	91	95	86	129	455	87,201

Appendix Table-32 Detailed Results EOUSA Matters OUT/EOUSA Cases IN: Links by Block

	Total Linked	**Linked by Block**	**% of Total Linked by Block**
block 1	1,159,278	1,159,278	99.43%
block 2	1,160,178	900	0.08%
block 3	1,165,873	5,695	0.49%

9. EOUSA Matters OUT/EOUSA Cases OUT

Appendix Table-33 Detailed Results EOUSA Matters OUT/EOUSA Cases OUT: Overall Link Rate

	% EOUSA Matters Out	% EOUSA Cases Out
1994	47.20%	37.50%
1995	44.38%	75.16%
1996	40.79%	82.30%
1997	54.96%	65.15%
1998	56.50%	74.03%
1999	56.24%	78.80%
2000	57.53%	80.40%
2001	56.42%	79.89%
2002	57.40%	80.96%
2003	56.64%	81.28%
2004	53.54%	87.11%
2005	53.17%	87.46%
2006	51.91%	85.17%
2007	50.00%	85.72%
2008	38.04%	87.83%
2009	17.54%	89.17%

Appendix Table-34 Detailed Results EOUSA Matters OUT/EOUSA Cases OUT: Links by EOUSA Cases OUT Year

									eoCOutYear								
eoMOutYear	**0**	**1994**	**1995**	**1996**	**1997**	**1998**	**1999**	**2000**	**2001**	**2002**	**2003**	**2004**	**2005**	**2006**	**2007**	**2008**	**2009**
0	0	36,900	14,388	10,717	22,598	17,842	15,896	15,583	16,286	16,132	16,556	10,750	10,824	13,544	12,562	11,247	10,320
1994	52,265	22,126	21,368	3,109	61	24	11	6	8	3	3	0	0	0	0	0	0
1995	57,686	14	22,171	23,351	278	143	29	19	9	3	3	0	0	0	0	0	0
1996	59,201	2	3	23,380	16,730	484	107	43	11	15	5	0	0	0	0	0	0
1997	45,489	0	0	0	25,164	24,187	3,853	1,028	380	243	171	134	89	100	60	54	35
1998	46,312	0	0	0	5	26,000	27,263	4,277	1,086	525	267	193	113	112	104	107	93
1999	50,450	0	0	0	8	4	27,797	29,387	4,639	1,495	554	285	188	159	118	130	86
2000	50,400	0	0	0	1	2	8	29,167	30,632	5,441	1,443	682	309	186	146	142	113
2001	52,396	0	0	0	2	0	0	4	27,934	31,916	5,032	1,463	665	338	227	131	126
2002	53,441	0	0	0	1	2	0	2	4	28,938	33,789	5,849	1,794	845	432	200	147
2003	56,422	0	0	0	0	1	1	1	1	5	30,587	33,346	6,360	1,910	858	355	282
2004	69,615	0	0	0	0	0	0	0	1	5	10	30,552	38,062	7,822	2,327	946	495
2005	67,946	0	0	0	0	0	0	0	1	4	6	67	27,740	37,998	7,853	2,406	1,064
2006	68,574	0	0	0	0	0	0	1	1	1	1	12	30	28,146	35,837	7,533	2,461
2007	72,706	0	0	0	0	1	0	0	0	0	1	8	22	55	27,340	37,586	7,690
2008	114,148	0	0	0	0	0	0	0	1	0	0	26	43	40	58	31,500	38,422
2009	160,653	0	0	0	0	0	0	0	0	2	0	33	59	58	46	62	33,917

This document is a research report submitted to the U.S. Department of Justice. This report has not
been published by the Department. Opinions or points of view expressed are those of the author(s)
and do not necessarily reflect the official position or policies of the U.S. Department of Justice.

Appendix Table-35 Detailed Results EOUSA Matters OUT/EOUSA Cases OUT: Links by EOUSA Matters OUT Year

									eoMOutYear								
eoCOutYear	**0**	**1994**	**1995**	**1996**	**1997**	**1998**	**1999**	**2000**	**2001**	**2002**	**2003**	**2004**	**2005**	**2006**	**2007**	**2008**	**2009**
0	0	52,265	57,686	59,201	45,489	46,312	50,450	50,400	52,396	53,441	56,422	69,615	67,946	68,574	72,706	114,148	160,653
1994	36,900	22,126	14	2	0	0	0	0	0	0	0	0	0	0	0	0	0
1995	14,388	21,368	22,171	3	0	0	0	0	0	0	0	0	0	0	0	0	0
1996	10,717	3,109	23,351	23,380	0	0	0	0	0	0	0	0	0	0	0	0	0
1997	22,598	61	278	16,730	25,164	5	8	1	2	1	0	0	0	0	0	0	0
1998	17,842	24	143	484	24,187	26,000	4	2	0	2	1	0	0	0	1	0	0
1999	15,896	11	29	107	3,853	27,263	27,797	8	0	0	1	0	0	0	0	0	0
2000	15,583	6	19	43	1,028	4,277	29,387	29,167	4	2	1	0	0	1	0	0	0
2001	16,286	8	9	11	380	1,086	4,639	30,632	27,934	4	1	1	1	1	0	1	0
2002	16,132	3	3	15	243	525	1,495	5,441	31,916	28,938	5	5	4	1	0	0	2
2003	16,556	3	3	5	171	267	554	1,443	5,032	33,789	30,587	10	6	1	1	0	0
2004	10,750	0	0	0	134	193	285	682	1,463	5,849	33,346	30,552	67	12	8	26	33
2005	10,824	0	0	0	89	113	188	309	665	1,794	6,360	38,062	27,740	30	22	43	59
2006	13,544	0	0	0	100	112	159	186	338	845	1,910	7,822	37,998	28,146	55	40	58
2007	12,562	0	0	0	60	104	118	146	227	432	858	2,327	7,853	35,837	27,340	58	46
2008	11,247	0	0	0	54	107	130	142	131	200	355	946	2,406	7,533	37,586	31,500	62
2009	10,320	0	0	0	35	93	86	113	126	147	282	495	1,064	2,461	7,690	38,422	33,917

Appendix Table-36 Detailed Results EOUSA Matters OUT/EOUSA Cases OUT: Links by Block

	Total Linked	**Linked by Block**	**% of Total Linked by Block**
block 1	975,253	975,253	97.12%
block 2	994,974	19,721	1.96%
block 3	1,004,184	9,210	0.92%

10. EOUSA Matters IN/ EOUSA Matters OUT

Appendix Table-37 Detailed Results EOUSA Matters IN/EOUSA Matters OUT: Overall Link Rate

	% EOUSA Matters Out	% EOUSA Matters In
1994	57.79%	87.46%
1995	79.36%	81.25%
1996	87.77%	71.78%
1997	72.67%	91.00%
1998	86.03%	92.98%
1999	89.37%	93.46%
2000	92.62%	93.48%
2001	93.20%	92.90%
2002	93.32%	93.71%
2003	93.53%	62.84%
2004	79.00%	95.30%
2005	86.21%	94.06%
2006	88.60%	92.53%
2007	89.70%	90.47%
2008	92.48%	88.84%
2009	92.65%	79.85%

Appendix Table-38 Detailed Results EOUSA Matters IN/EOUSA Matters OUT: Links by EOUSA Matters OUT Year

	eoMOutYear															
eoMInYear	**0**	**1994**	**1995**	**1996**	**1997**	**1998**	**1999**	**2000**	**2001**	**2002**	**2003**	**2004**	**2005**	**2006**	**2007**	**2008**	**2009**
0	0	41,785	21,406	12,224	27,603	14,876	12,256	8,761	8,180	8,384	8,417	31,464	20,004	16,254	14,970	13,856	14,323
1994	12,446	56,520	22,208	7,295	231	138	116	89	72	60	76	0	0	0	0	0	0
1995	19,170	165	59,823	21,676	361	262	218	166	124	136	119	0	0	0	0	0	0
1996	27,589	198	111	58,616	9,543	547	343	278	217	179	155	0	0	0	0	0	0
1997	9,902	62	22	34	63,077	21,354	7,546	3,503	2,040	1,219	684	249	120	99	54	32	37
1998	8,119	46	22	30	38	69,237	22,149	7,537	3,883	2,373	1,159	556	219	168	73	53	30
1999	7,722	56	23	27	39	15	72,538	21,972	7,320	4,278	2,075	1,009	460	206	103	79	72
2000	8,053	53	35	29	38	9	54	76,267	22,515	8,231	4,254	2,086	916	508	241	174	96
2001	8,646	57	31	19	21	11	50	63	75,829	22,288	7,604	3,505	1,715	996	495	330	158
2002	7,821	42	25	30	35	8	29	35	53	78,291	23,894	6,977	3,214	1,991	985	541	364
2003	48,342	0	0	1	1	0	1	1	1	5	81,683	40	1	1	0	1	0
2004	6,638	0	0	0	0	0	0	0	0	0	2	102,537	18,177	7,250	3,602	1,905	1,104
2005	8,171	0	0	0	0	0	0	0	0	0	0	147	99,384	17,612	6,874	3,398	2,004
2006	10,001	0	0	0	0	0	0	0	0	0	0	153	96	96,924	16,613	6,385	3,763
2007	13,195	0	0	0	0	0	0	0	0	0	4	243	155	129	101,003	17,002	6,679
2008	19,920	0	0	0	0	0	0	0	0	0	0	387	258	180	161	140,205	17,459
2009	37,958	0	0	0	0	0	0	0	0	0	3	482	366	279	235	277	148,741

Appendix Table-39 Detailed Results EOUSA Matters IN/EOUSA Matters OUT: Links by EOUSA Matters IN Year

									eoMInYear								
eoMOutYear	**0**	**1994**	**1995**	**1996**	**1997**	**1998**	**1999**	**2000**	**2001**	**2002**	**2003**	**2004**	**2005**	**2006**	**2007**	**2008**	**2009**
0	0	12,446	19,170	27,589	9,902	8,119	7,722	8,053	8,646	7,821	48,342	6,638	8,171	10,001	13,195	19,920	37,958
1994	41,785	56,520	165	198	62	46	56	53	57	42	0	0	0	0	0	0	0
1995	21,406	22,208	59,823	111	22	22	23	35	31	25	0	0	0	0	0	0	0
1996	12,224	7,295	21,676	58,616	34	30	27	29	19	30	1	0	0	0	0	0	0
1997	27,603	231	361	9,543	63,077	38	39	38	21	35	1	0	0	0	0	0	0
1998	14,876	138	262	547	21,354	69,237	15	9	11	8	0	0	0	0	0	0	0
1999	12,256	116	218	343	7,546	22,149	72,538	54	50	29	1	0	0	0	0	0	0
2000	8,761	89	166	278	3,503	7,537	21,972	76,267	63	35	1	0	0	0	0	0	0
2001	8,180	72	124	217	2,040	3,883	7,320	22,515	75,829	53	1	0	0	0	0	0	0
2002	8,384	60	136	179	1,219	2,373	4,278	8,231	22,288	78,291	5	0	0	0	0	0	0
2003	8,417	76	119	155	684	1,159	2,075	4,254	7,604	23,894	81,683	2	0	0	4	0	3
2004	31,464	0	0	0	249	556	1,009	2,086	3,505	6,977	40	102,537	147	153	243	387	482
2005	20,004	0	0	0	120	219	460	916	1,715	3,214	1	18,177	99,384	96	155	258	366
2006	16,254	0	0	0	99	168	206	508	996	1,991	1	7,250	17,612	96,924	129	180	279
2007	14,970	0	0	0	54	73	103	241	495	985	0	3,602	6,874	16,613	101,003	161	235
2008	13,856	0	0	0	32	53	79	174	330	541	1	1,905	3,398	6,385	17,002	140,205	277
2009	14,323	0	0	0	37	30	72	96	158	364	0	1,104	2,004	3,763	6,679	17,459	148,741

Appendix Table-40 Detailed Results EOUSA Matters IN/EOUSA Matters OUT: Links by Block

	Total Linked	**Linked by Block**	**% of Total Linked by Block**
block 1	1,783,962	1,783,962	98.72%
block 2	1,797,555	13,593	0.75%
block 3	1,807,125	9,570	0.53%

C. Detailed Comparison to First Generation System

1. EOUSA Matters OUT/USMS IN

When comparing to the previous linking system, we consider the links made to a particular USMS observation.

Further, we only consider USMS observations that appear in both the first generation system and the new system. Because the USMS observations considered by the dyad system are screened to remove supervision violators and material witnesses and the first generation system does not screen USMS observations, there are some cases where a USMS observation was linked in the old system, but screened out of the new.

Also note that there are observations that are not included in the first generation universe, but are included in the new dyad system. This is primarily because the value of court case number in the USMS IN SAF had not been set. If this value was missing or unknown, then the first generation system was unable to make a link. Of the approximate 150,000 observations that were in the dyad system input (for years 1994-2005) and absent from the first generation system, over 90% have an unknown value indicated for court case number. The dyad system however, is able to make use of other information on the record in the linking process.

Of the approximately 158,000 USMS observations that are found in the first generation system only and screened out of the dyad system universe because they were supervision violators or material witnesses, about 13,600 were linked to an EOUSA record. These links will not be reproduced by the dyad system.

A comparison of the linking results for USMS observations common to both systems is as follows:

Appendix Table-41Comparison of Links Made to USMS Observations in First Generation Link System and Dyad Link System USMS Observations Years 1994-2005

	Frequency	Percent
First Gen. System and Dyad links are identical	497,322	51.66
First Gen. System has no link, Dyad does	226,875	23.57
First Gen. System has link, Dyad does not	19,278	2.00
First Gen. System and Dyad links are different	21,765	2.26
First Gen. System and Dyad are unlinked	197,376	20.50

Appendix Table-42 Year of USMS Observation Where the First Generation System Has No Link and Dyad System Does

	Frequency	Percent
1994	11,076	4.88
1995	8,329	3.67
1996	7,179	3.16
1997	9,033	3.98
1998	10,581	4.66
1999	10,370	4.57
2000	12,381	5.46
2001	11,070	4.88
2002	24,035	10.59
2003	53,001	23.36
2004	36,553	16.11
2005	33,268	14.66

The majority of the new dyad system links are in the later years, these USMS records are in large part being linked to EOUSA Matters OUT observations in FY2006-2008. The increase in linked observations in 2003 is a result of a data correction to the EOUSA data that was not in this version of the first generation linking system.

2. AOUSC IN/EOUSA IN

When comparing to the previous linking system, we consider the links made to a particular AOUSC Cases IN observation.

Results are as follows:

Appendix Table-43 Comparison of Links Made to AOUSC IN Observations in First Generation Link System and Dyad System AOUSC Years 1994-2005

	Frequency	Percent
First Gen. System and Dyad links are identical	564,746	59.00
First Gen. System has no link, Dyad does	183,885	19.21
First Gen. System has link, Dyad does not	17,801	1.86
First Gen. System and Dyad links are different	5,814	0.61
First Gen. System and Dyad are unlinked	184,904	19.32

Appendix Table-44 Year of AOUSC IN Observation Where the First Generation System Has No Link and Dyad System Does

	Frequency	Percent
1994	6,267	3.41
1995	5,635	3.06
1996	5,344	2.91
1997	6,816	3.71
1998	9,810	5.33
1999	9,255	5.03
2000	9,431	5.13
2001	9,337	5.08
2002	10,155	5.52
2003	49,738	27.05
2004	15,317	8.33
2005	46,780	25.44

As with the EOUSA-USMS links, the AOUSC observations in 2005 have many more links because they are linked to EOUSA observations in FY2006-2009. The increase in linked observations in 2003 is a result of a data correction to the EOUSA data that was not in this version of the first generation linking system.

Appendix Table-45 Comparison of Link Rates for AOUSC IN and EOUSA IN

	Dyad		First Gen. System		Change in Link Rate	
	% AOUSC Linked	% EOUSA Linked	% AOUSC Linked	%EOUSA Linked	% AOUSC Linked	% EOUSA Linked
1994	74.25%	76.51%	66.26%	68.64%	7.99%	7.87%
1995	79.27%	78.23%	72.43%	72.38%	6.84%	5.85%
1996	78.51%	79.26%	73.19%	73.53%	5.32%	5.72%
1997	75.16%	75.47%	70.77%	67.40%	4.40%	8.06%
1998	76.97%	87.38%	67.58%	75.67%	9.39%	11.70%
1999	79.24%	87.67%	71.11%	77.05%	8.13%	10.63%
2000	80.95%	86.90%	73.19%	77.14%	7.77%	9.76%
2001	81.03%	87.07%	73.42%	77.48%	7.61%	9.60%
2002	80.40%	86.49%	73.04%	77.05%	7.35%	9.44%
2003	78.87%	87.98%	26.33%	28.38%	52.55%	59.60%
2004	80.10%	80.99%	66.52%	66.69%	13.58%	14.30%
2005	79.04%	76.86%	29.17%	30.08%	49.87%	46.78%
2006	80.22%	76.65%	NA	NA	NA	NA
2007	80.31%	79.33%	NA	NA	NA	NA
2008	82.64%	80.56%	NA	NA	NA	NA
2009	81.60%	79.98%	NA	NA	NA	NA

3. AOUSC OUT/EOUSA OUT

When comparing to the previous linking system, we will consider the links made to a particular AOUSC Cases OUT observation.

Results are as follows.

Appendix Table-46 Comparison of Links Made to AOUSC OUT Observations in First Generation Link System and Dyad System AOUSC Year 1994-2005

	Frequency	Percent
First Gen. and Dyad links are identical	553,372	62.1
First Gen. has no link, Dyad does	147,589	16.56
First Gen. has link, Dyad does not	41,518	4.66
First Gen. and Dyad links are different	28,520	3.2
First Gen. and Dyad are unlinked	120,063	13.47

Appendix Table-47 Year of AOUSC OUT Observation Where the First Generation System Has No Link and Dyad System Does

	Frequency	Percent
1994	28,330	19.27
1995	8,592	5.84
1996	5,845	3.98
1997	5,643	3.84
1998	8,273	5.63
1999	8,471	5.76
2000	8,551	5.82
2001	7,941	5.33
2002	8,325	5.66
2003	12,489	8.5
2004	34,733	23.63
2005	69,916	6.75

The increase in linked observations in 2004 is a result of a data correction to the EOUSA data that was not in this version of the first generation system.

The higher number of AOUSC-EOUSA OUT links missed by the dyad system but made in the first generation system (relative to other dyads) can be explained by several factors. First, there are a set of observations where blocking variables don't line up, and so the dyad system will never compare them (that is, dates which are used in blocking do not line up). Second, in the first generation system, there are some EOUSA Matters OUT observations that are non-magistrate matters with a link to an AOUSC OUT observation. These EOUSA Matters OUT observations will not be in the EOUSA file used by the dyad system.[12] The percentage of links missed by the dyad system is fairly constant across all years.

Appendix Table-48 Comparison of Link Rates for AOUSC OUT and EOUSA Cases OUT

	Dyad		First Gen. System		Change in Link Rate	
	% AOUSC Linked	**% EOUSA Linked**	**% AOUSC Linked**	**%EOUSA Linked**	**% AOUSC Linked**	**% EOUSA Linked**
1994	74.26%	76.61%	29.35%	31.37%	44.91%	45.24%
1995	78.23%	75.86%	63.73%	63.68%	14.50%	12.17%
1996	79.56%	79.66%	70.25%	72.80%	9.31%	6.86%
1997	79.69%	77.22%	70.66%	72.67%	9.03%	4.55%
1998	78.00%	76.71%	65.69%	71.16%	12.31%	5.56%
1999	79.30%	77.83%	67.23%	71.82%	12.07%	6.01%
2000	82.81%	77.47%	71.00%	72.49%	11.82%	4.98%
2001	82.89%	76.38%	71.89%	73.07%	11.00%	3.31%
2002	85.16%	78.00%	73.91%	72.44%	11.26%	5.56%
2003	83.54%	77.52%	63.18%	69.83%	20.36%	7.70%
2004	83.50%	80.69%	32.43%	32.44%	51.07%	48.24%
2005	82.52%	80.53%	67.43%	69.39%	15.09%	11.14%
2006	83.70%	78.04%	NA	NA	NA	NA
2007	81.88%	78.79%	NA	NA	NA	NA
2008	84.32%	80.94%	NA	NA	NA	NA
2009	85.12%	81.48%	NA	NA	NA	NA

NOTES: EOUSA observations shown are Cases OUT only

There appears to be considerable noise in the first generation system links. This is likely due to the fact that the EOUSA Matters OUT observation is the base of the EOUSA linkages in the first generation linking system, whereas we are comparing Cases OUT observations.

4. AOUSC OUT/USSC OUT

When comparing to the previous linking system, we consider the links made to a particular AOUSC Cases OUT observation. Further, we only consider AOUSC observations that appear in both the first generation system and the dyad system. Because the AOUSC observations considered by the dyad system are screened to identify those who were convicted, and the first generation observations are not, there are some cases where an AOUSC observation was linked in the first generation system, but screened out in the dyad system.

Of the approximately 146,000 AOUSC observations that are found in the first generation system only and screened out of the dyad system universe, about 2,300 were linked to a USSC record. These links will not be reproduced by the dyad system. Further any links made to sealed records in the first generation system will not be made in the dyad system.

Results are as follows:

[12] Recall the dyad system only considers magistrate matters from the EOUSA Matters OUT file when linking to AOUSC OUT.

Appendix Table-49 Comparison of Links Made to AOUSC OUT Observations in First Generation Link System and Dyad System AOUSC Year 1994-2005

	Frequency	Percent
First Gen. and Dyad links are identical	572,725	64.27
First Gen. has no link, Dyad does	42,665	4.79
First Gen. has link, Dyad does not	29,266	3.28
First Gen. and Dyad links are different	2,839	0.32
First Gen. and Dyad are unlinked	243,537	27.33

Appendix Table-50 Year of AOUSC OUT Observation Where the First Generation System Has No Link and Dyad System Does

	Frequency	Percent
1994	2,946	6.47
1995	2,705	5.94
1996	2,798	6.15
1997	3,421	7.52
1998	3,626	7.97
1999	3,618	7.95
2000	4,085	8.98
2001	4,111	9.03
2002	4,019	8.83
2003	4,469	9.82
2004	5,228	11.49
2005	4,485	9.85

Appendix Table-51 Comparison of Link Rates for AOUSC OUT and USSC OUT

	Dyad		First Gen. System		change in link rate	
	% AOUSC Linked	% USSC Linked	% AOUSC Linked	% USSC Linked	% AOUSC Linked	% USSC Linked
1994	58.13%	91.24%	57.05%	90.00%	1.09%	1.24%
1995	59.87%	89.71%	58.80%	88.01%	1.06%	1.69%
1996	62.97%	92.12%	62.42%	90.71%	0.56%	1.41%
1997	68.85%	92.76%	67.60%	90.09%	1.25%	2.67%
1998	65.55%	91.08%	64.85%	89.72%	0.69%	1.35%
1999	66.56%	91.65%	65.55%	90.73%	1.02%	0.92%
2000	70.51%	91.48%	68.83%	89.85%	1.68%	1.63%
2001	70.69%	91.83%	69.26%	89.93%	1.43%	1.90%
2002	73.30%	92.38%	71.97%	91.20%	1.32%	1.18%
2003	75.46%	92.29%	74.47%	90.98%	0.99%	1.31%
2004	76.38%	91.69%	75.04%	89.82%	1.34%	1.87%
2005	75.97%	91.52%	76.15%	89.96%	-0.18%	1.56%
2006	74.75%	91.36%	NA	NA	NA	NA
2007	76.16%	92.98%	NA	NA	NA	NA
2008	77.69%	93.62%	NA	NA	NA	NA
2009	79.05%	93.53%	NA	NA	NA	NA

NOTE: AOUSC columns show all AOUSC Observations, not just those with outcome in (1,2,3,4)

The increase in link rate when using the dyad system for this dyad is modest.

5. BOP IN/USSC OUT

When comparing to the previous linking system, we consider the links made to a particular USSC observation.

Further, we only consider USSC observations that appear in both the first generation system and the dyad system. Because the USSC observations considered by the new system are screened to identify those who were sentenced to prison, and the first generation system observations are not, there are some cases where a USSC observation was linked in the first generation system, but screened out in the dyad system.

Of the approximate 113,000 USSC observations that are found in the first generation system only and screened out of the dyad universe, about 15,000 were linked to a BOP record. These links will not be reproduced by the new system. Further, approximately 41,000 USSC observations in the first generation system that are also present in the new system were linked in the first generation system to BOP IN observations that are now being screened out. These links will also not be reproduced by the dyad system.

A comparison of the linking results for USSC observations common to both systems is as follows:

Appendix Table-52 Comparison of Links Made to USSC OUT Observations in First Generation Link System and Dyad System USSC Year 1994-2005

	Frequency	Percent
First Gen. and Dyad links are identical	246,607	36.76
First Gen. has no link, Dyad does	191,695	28.57
First Gen. has link, Dyad does not	33,013	4.92
First Gen. and Dyad links are different	34,001	5.07
First Gen. and Dyad are unlinked	165,624	24.69

Appendix Table-53 Year of USSC OUT Observation Where the First Generation System Has No Link and Dyad System Does

	Frequency	Percent
1994	4,722	2.46
1995	5,219	2.72
1996	5,995	3.13
1997	6,978	3.64
1998	8,516	4.44
1999	9,657	5.04
2000	11,606	6.05
2001	13,784	7.19
2002	17,957	9.37
2003	25,182	13.14
2004	34,887	18.2
2005	47,192	24.62

This dyad has a higher percentage of links that are different across systems. This is mostly a result of the screening rules that are present in the dyad system and not in the first generation system. Additionally, as with other dyads there are more links made with the new system than the first generation system in the later years because a portion of the links to USSC observations in 2003 – 2005 are to BOP records in 2006-2008 and thus not included in the first generation system.