

**Education Research & Data Center (ERDC)**  
**EFFECTIVE LINKING OF LONGITUDINAL EDUCATION DATA**  
**USING LINK PLUS AND THE LINK PLUS TOOLKIT**

**NCES 25<sup>th</sup> Annual STATS-DC 2012 Data Conference**

Washington DC

July 13, 2012

# About the ERDC

# VISION, MISSION

## Washington State Education Research & Data Center

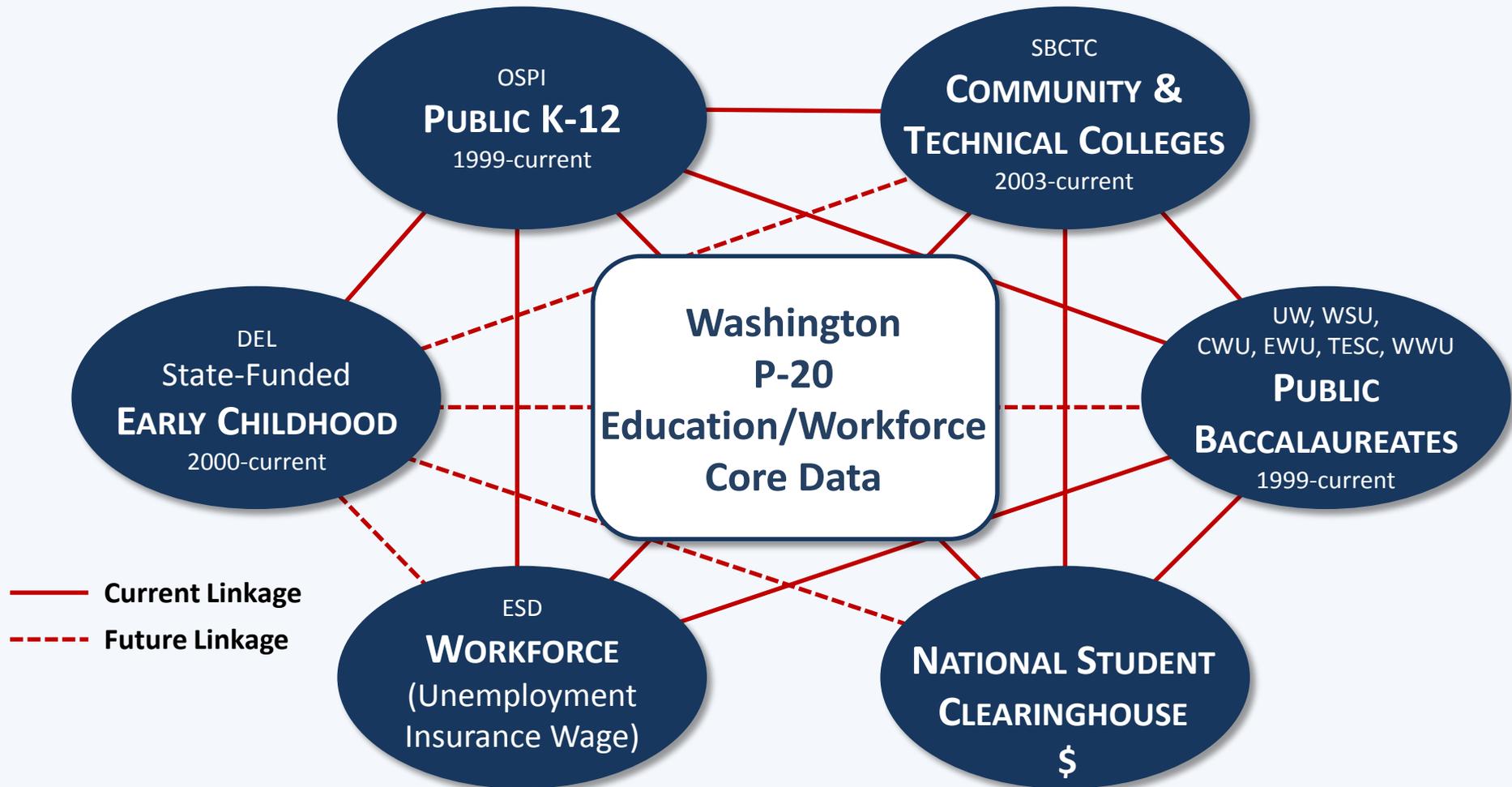
### **Vision**

To promote a seamless, coordinated preschool-to-career (P-20W) experience for all Washington State learners by providing objective analysis and information.

### **Mission**

To develop longitudinal information spanning the P-20W system in order to facilitate analyses, provide meaningful reports, collaborate on education research, and share data.

# P-20 LONGITUDINAL DATA ELEMENTS



OSPI = Office of Superintendent of Public Instruction  
 DEL = Department of Early Learning  
 ESD = Employment Security Department  
 SBCTC = State Board for Community & Technical Colleges

UW = University of Washington  
 WSU = Washington State University  
 CWU = Central Washington University

EWU = Eastern Washington University  
 TESC = The Evergreen State College  
 WWU = Western Washington University

# **The Project: Linking Divorce Data to Death Data**

# WA State Male Divorce Data\*

CERTNUM	HSB_LAST	HSB_FRST	HSB_MI	hsb_dob	HSB_RES	mar_date	dc_date	DC_TYPE
1995020828	HAYENGA	RICHARD	W	12/13/68	3105	08/07/91	09/06/95	1
1995020831	RASK	TRACY	C	12/02/68	3109	04/01/92	09/06/95	1
1995020832	TAYLOR	MICHAEL	L	03/22/69	3100	06/10/88	09/06/95	1
1995020849	CROOK	CARY	A	11/10/66	1711	08/18/90	09/11/95	1
1995020858	SCHAMING	DARRELL	D	06/16/62	3103	08/05/89	09/13/95	1
1995020861	RUSSELL	KEVIE	L	08/19/67	3102	07/01/89	09/13/95	1
1995020862	REESE	MARK	A	10/15/70	3102	04/27/91	09/13/95	1
1995020869	MILLER	GLEN	P	12/16/68	3101	11/04/87	09/13/95	1
1995020870	CARDOZA	ROBERT	F	11/27/64	3105	09/29/90	09/13/95	1
1995020877	VIDATO	KEVIN	W	08/21/64	1705	05/22/87	09/13/95	1
1995020878	WARREN	NORMAN	D	10/28/68	3104	08/14/93	09/13/95	1
1995020881	AUSTIN	RICHARD	P	07/21/67	1500	06/29/91	09/13/95	1
1995020882	HUTSON	PATRICK	E	03/09/64	8899	10/10/92	09/13/95	1
1995020884	BURGESS	ARTHUR	B	05/31/66	1701	12/23/93	09/13/95	3
1995020885	BRYANT	JAMES	O	10/31/63	3102	07/04/93	09/13/95	1
1995020887	KUTSCH	TODD	M	03/31/70	3103	02/06/93	09/14/95	1
1995020888	SHACKLES	TIMMY	A	03/15/65	3100	11/23/89	09/14/95	1

\* Washington State divorce data is public disclosable data.

# WA State Male Death Certificate Data\*

certno	LastName	FirstName	MiddleName	DOB	city_res	SSN	dth_date	age	ecode	underly*
1995022316	ANTON	WALTER	FOSTER	01/24/36	1701	546-46-4899	07/29/95	59		4292
1995022317	PUTNAM	OTHA	GEORGE	12/18/59	1719	542-68-3907	07/29/95	35	E	E850
1995022318	HALL	KELLY	DUFFY	05/31/78	1704	524-67-5305	07/29/95	17	E	E815
1995022320	SWEENEY	HOWARD	TED	03/28/34	1701	531-30-3457	07/29/95	61		4292
1995022321	RILEY	JOHN	FRANCIS	06/05/22	1500	538-07-7225	07/29/95	73		436
1995022322	LEE	DUK	KYU	05/15/23	1719	531-86-0139	07/29/95	72		4140
1995022323	FIELD	JOHN	CHRISTOPHER	12/07/32	1700	546-60-3029	07/29/95	62		4292
1995022325	WELFELT	RICHARD	DECKER	12/14/27	1701	531-20-7886	07/29/95	67		2503
1995022328	USMIAL	JOHN	STEPHEN	06/20/30	1701	384-28-0883	07/29/95	65		5712
1995022329	MCNULTY	WILLIAM		03/19/15	1701	538-10-5280	07/29/95	80		5070
1995022330	ERLEY	ROBERT	DONALD	05/18/21	1700	533-12-7208	07/29/95	74		1541
1995022331	BARTZ	ALBERT	FREDRICK	09/28/05	1800	469-03-4298	07/29/95	89		431
1995022332	HUNTER	WILLIAM	MICHAEL	03/07/41	1701	560-54-1650	07/29/95	54		4149
1995022335	DOLHANYK	JAMES	ALAN	08/20/60	3100	538-68-1877	07/29/95	34		1724
1995022337	HALL	SIGVARD	S	01/22/05	1701	539-42-2497	07/29/95	90		410
1995022342	WOOD	JAMES	DAVID	06/28/38	1500	531-36-3875	07/29/95	57		2028
1995022343	ENSLEY	SAMUEL	F	05/11/02	1701	414-05-2034	07/29/95	93		4149
1995022344	WALKER	REGINALD	JOSPEH	08/29/58	1701	535-60-6474	07/29/95	36	E	E858

\*Cause-of-death data used ICD-9 codes from 1980 to 1998, and ICD-10 codes from 1999 to present.

\*Death certificate data is public disclosable data.

# Walter Foster Anton's SSN

Decoding Social Security Numbers in One Step - Mozilla Firefox

File Edit View History Bookmarks Tools Help

Step Decoding Social Security Numbers in On... +

Step stevemorse.org/ssn/ssn.html

## Five-Digit Decoder

**SSN starting with** 546 - 46 - XXXX

**was issued in** 546-46: California, 1952

*Where and when card was issued, not where and when person was born*

# Why Probabilistic Linking?

# Identifiers in the Perfect World

```
SELECT K12.*, College.*  
FROM K12  
INNER JOIN College  
    ON K12.Bulletproof_Surefire_Global_Student_ID =  
        College.Bulletproof_Surefire_Global_Student_ID
```

# Identifiers in the Other Perfect World

```
SELECT K12.*, College.*  
FROM K12  
INNER JOIN College  
    ON K12.SSN = College.SSN
```

Note: Every student has a valid, properly assigned SSN.

# Potential Matched Pair in Our Imperfect World

<b>LastName</b>	<b>FirstName</b>	<b>DOB</b>	<b>SSN</b>
Public	John	<b>7/4/1976</b>	<b>555-55-5555</b>
Public	John	<b>4/7/1976</b>	<b>555-55-5556</b>

# Link Plus and its Limitations

# About Link Plus: Internet Resources

## Overview of Link Plus\*

- [www.nri-inc.org/projects/OSA/LinkPlusOverviewMarch2007.pdf](http://www.nri-inc.org/projects/OSA/LinkPlusOverviewMarch2007.pdf)

## An excellent Link Plus manual\*

- [www.npaihb.org/images/epicenter\\_docs/NW-Idea/Link Plus manual 202011 FINAL.docx](http://www.npaihb.org/images/epicenter_docs/NW-Idea/Link%20Plus%20manual%202011%20FINAL.docx)

\* Links accessed on July 7, 2012

# Link Plus Configuration Screen

File 1:

File 2:

**Select blocking variables**

Data Item (File 1)	Data Item (File 2)	Phonetic System
LastName	HSB_LST	Soundex
FirstName	HSB_FRST	Soundex
dob	hsb_dob	[None]

**Select ID variables (File 1)**

**Select ID variables (File 2)**

**Select matching variables and methods**

Data Item (File 1)	Data Item (File 2)	Matching Method
LastName	HSB_LST	Last Name
FirstName	HSB_FRST	First Name
MiddleName	HSB_MI	Middle Name
dob	hsb_dob	Date
city_res	HSB_RES	Exact

Direct Method

Cutoff Value:

Results will be saved to

Missing Value (File 1)  Missing Value (File 2)

Please configure your linkage, and then click the Run button.

# Manual Review of Uncertain Matches\*

Link Plus  
Manual Review Data Tools Help

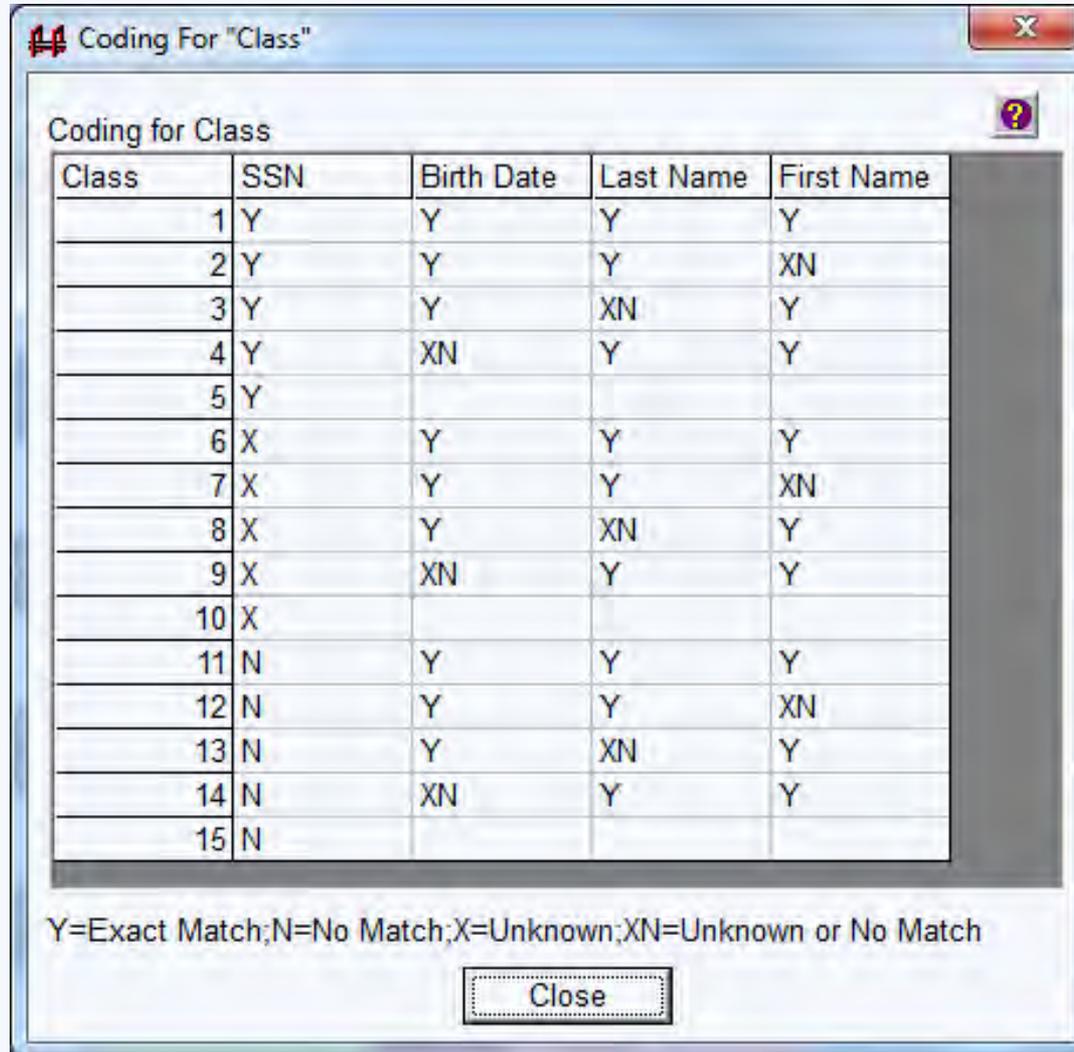
Linkage Report=L:\Report\Unmodified\DeathToDivorce.txt

= true matches     = false matches     = uncertain matches     = unmatched values     = missing values

Score	Class	certno_File1	CERTNUM_File2	LastName;HSB_LS	FirstName;HSB_FR	MiddleName;HSB_I	dob;hsb_dob	city_res;HSB_RE
17.4	9		2000023703	OLMOS	RAUL	A	11181965	3303
	9	1995031696		BISHOPRICK	STANLEY		05021904	0600
17.4	9		2006002501	BISHOPRICK	STANLEY	J	04121967	0600
	7	2009005535		JOHNSON	MATTHEW	CARL	11211964	3201
17.4	7		1992001779	JOHNSON	MATTHE	C	11211964	3200
	9	1982000412		MOREAU	JOSEPH	M	05101906	1701
17.4	9		1994020216	MOREAU	JOSEPH	M	06081963	1701
	7	2010006193		DAVIDSON	DANIEL	CHRISTIAN	08061965	1703
17.4	7		1992016663	DAVIDSON	DAN	C	08061965	2702
	9	2001038520		DIZARD	JOSEPH	LEWIS	06081925	3100
17.3	9		2001026790	DIZARD	JOSEPH	L	05101970	3100
	9	2008069414		COLLICOTT	DENNY	LAWRENCE	05051935	1305
17.3	9		2006015497	COLLICOTT	DENNY	L	05251967	3907
	9	1994004057		GREGOIRE	RAYMOND	ALBERT	06031922	2500
17.3	9		1990026825	GREGOIRE	RAYMOND	A	04211964	2500
	6	1997028343		JONES	JAMES	DUVAL	10011971	3403
17.3	6		1993025840	JONES	JAMES	D	10011971	3404
	7	2007065602		JONES	BRENTON	RUPERT	02261969	0402
17.3	7		1998014647	JONES	BRANTON	R	02261969	0400
	8	1994014777		JIM NANAMKIN	JERRY	JOSEPH	10281969	2400
17.3	8		1993019071	NANAMKIN	JERRY	J	10281969	2400
	8	2003020660		WESSINGER	NOAH	FREDERICK	07201933	2000

\* Screenshot contains only public, disclosable data.

# Hardcoded Class Definitions in Link Plus



Coding For "Class"

Coding for Class

Class	SSN	Birth Date	Last Name	First Name
1	Y	Y	Y	Y
2	Y	Y	Y	XN
3	Y	Y	XN	Y
4	Y	XN	Y	Y
5	Y			
6	X	Y	Y	Y
7	X	Y	Y	XN
8	X	Y	XN	Y
9	X	XN	Y	Y
10	X			
11	N	Y	Y	Y
12	N	Y	Y	XN
13	N	Y	XN	Y
14	N	XN	Y	Y
15	N			

Y=Exact Match;N=No Match;X=Unknown;XN=Unknown or No Match

Close

# Island Problem: Good Pairs Among Bad Pairs\*

Manual Review Data Tools Help

= true matches     = false matches     = uncertain matches

Score	Class	LastName;HSB_LS	FirstName;HSB_FR	MiddleName;HSB_I	dob;hsb_dob	city_res;HSB_RE
	9	BOSTON	MONTA	E	01161935	2701
<input type="checkbox"/> 17.7	9	BOSTON	MONTA	E	00251965	1701
	9	CLARK	MARK	DAVID	07241962	3201
<input type="checkbox"/> 17.7	9	CLARK	MARK	D	07241963	3201
	9	CLARK	MARK	DAVID	07241962	3201
<input type="checkbox"/> 17.7	9	CLARK	MARK	D	07241963	3201
	9	WEBER	JONATHAN	S	03161905	1702
<input type="checkbox"/> 17.7	9	WEBER	JONATHAN	S	05211971	1702
	9	TRAN	TUAN		12311925	2701
<input type="checkbox"/> 17.7	9	TRAN	TUAN	A	06161971	2701
	9	RATTAZZI	LUCA	CRISTIANO	10121937	1707
<input type="checkbox"/> 17.7	9	RATTAZZI	LUCA		07131968	1701
	9	BRINCEFIELD	JEFFREY	TROY	03111969	0900
<input type="checkbox"/> 17.7	9	BRINCEFIELD	JEFFREY	T	03311969	0901
	9	SUTTLES	WALTER	RANDOLPH	07311949	2600
<input type="checkbox"/> 17.7	9	SUTTLES	WALTER	R	08191969	2600
	9	COOPER	FRANK	JOHN	01061920	2100

\* Screenshot contains only public, disclosable data.

# Summary of Link Plus Limitations

- **You can accept and reject matches by score only.**
- **Link Plus Classes are hard coded.**
- **Island problem, good matches surrounded by bad matches and bad matches surrounded by good matches.**
- **All potential matches must always go through the manual review function every time.**

# **Overcoming Link Plus's Limitations: The Link Plus Toolkit**

# Link Plus Toolkit

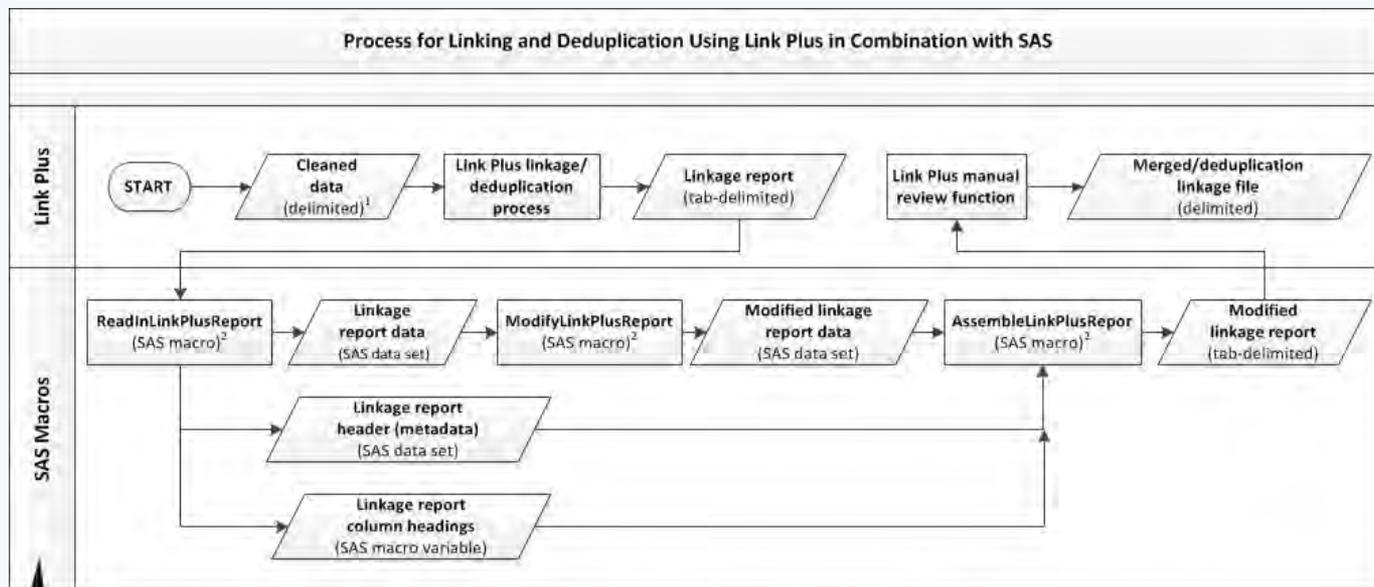
- **A set of SAS macros that add functionality to Link Plus**
- **Available on the ERDC website here:**  
**[www.erdcc.wa.gov/briefs/technical/](http://www.erdcc.wa.gov/briefs/technical/)**

# Link Plus Toolkit Contents

The Link Plus Toolkit currently contains six SAS macros:

- **ReadInLinkPlusReport.sas**
- **ModifyLinkPlusReport.sas**
- **SeparateByCondition.sas**
- **AssembleLinkPlusReport.sas**
- **CreateLinkPlusExportFile.sas**
- **MergeLinkPlusLinkageFiles.sas**

# Reading in Link Plus Linkage Report into SAS



- 1: Data imported into Link Plus can be either delimited (i.e. comma, tab, custom) or in a fixed field format.
- 2: The SAS macros ReadInLinkPlusReport.sas, ModifyLinkPlusReport.sas and AssembleLinkPlusReport.sas are available for download at the ERDC web page [www.erdc.wa.gov/briefs/technical](http://www.erdc.wa.gov/briefs/technical)

## Example Usage of SAS Macros

```

%ReadInLinkPlusReport(LinkRptPathIn=S:\Report\LinkageReport.txt, MetadataDsnOut=Metadata,
ReportDsnOut=ReportData)

%ModifyLinkPlusReport(ReportDsnIn=ReportData, ReportDsnOut=ReportData_Modified)

proc freq data=ReportData_Modified;
  title "ReportData Modified";
  table class / nopercnt;
run;

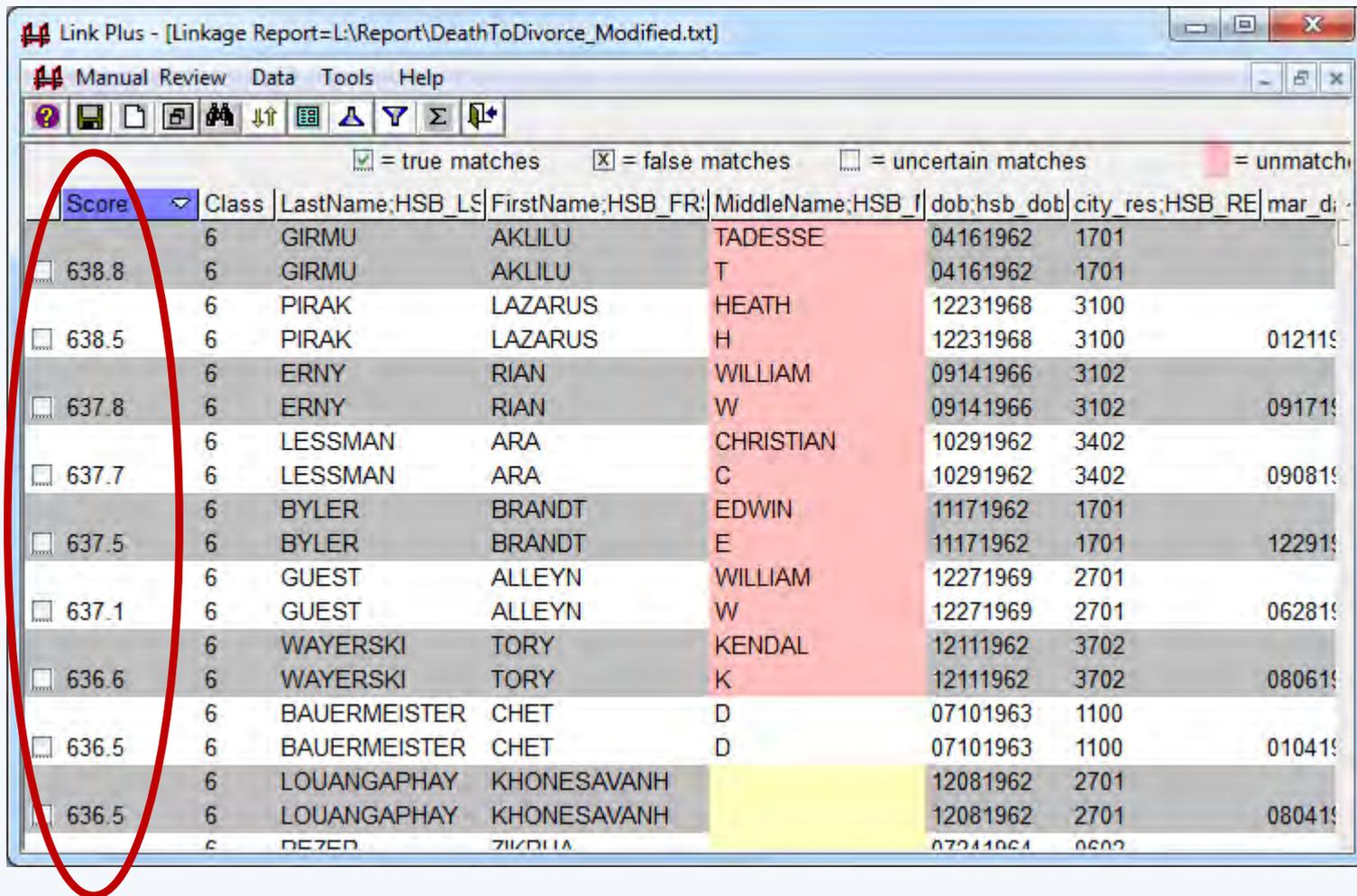
%AssembleLinkPlusReport(MetadataDsnIn=Metadata, ReportDsnIn=ReportData_Modified, LinkRptPathOut=S:\
Report\ReportData_Modified.txt)
  
```

# “Guts” of ModifyLinkPlusReport.sas Macro

```
data &ReportDsnOut.;  
  set &ReportDsnIn.;  
  
  orig_score = score;  
  score = (class * 100) + score;  
run;
```

- Results in non-overlapping range of reassigned scores for each class.

# ModifyLinkPlusReport.sas Effects on Score



The screenshot shows a SAS window titled "Link Plus - [Linkage Report=L:\Report\DeathToDivorce\_Modified.txt]". The window contains a table with the following columns: Score, Class, LastName;HSB\_LS, FirstName;HSB\_FR, MiddleName;HSB\_I, dob;hsb\_dob, city\_res;HSB\_RE, and mar\_d. The 'Score' column is highlighted with a red circle. The table contains 18 rows of data, with the first two rows having a score of 638.8, the next two 638.5, the next two 637.8, the next two 637.7, the next two 637.5, the next two 637.1, the next two 636.6, the next two 636.5, and the last row 636.5. The 'MiddleName;HSB\_I' column is highlighted in red for the first 17 rows, and the 'MiddleName;HSB\_I' column is highlighted in yellow for the last row.

Score	Class	LastName;HSB_LS	FirstName;HSB_FR	MiddleName;HSB_I	dob;hsb_dob	city_res;HSB_RE	mar_d
638.8	6	GIRMU	AKLILU	TADESSE	04161962	1701	
638.8	6	GIRMU	AKLILU	T	04161962	1701	
638.5	6	PIRAK	LAZARUS	HEATH	12231968	3100	
638.5	6	PIRAK	LAZARUS	H	12231968	3100	012119
637.8	6	ERNY	RIAN	WILLIAM	09141966	3102	
637.8	6	ERNY	RIAN	W	09141966	3102	091719
637.7	6	LESSMAN	ARA	CHRISTIAN	10291962	3402	
637.7	6	LESSMAN	ARA	C	10291962	3402	090819
637.5	6	BYLER	BRANDT	EDWIN	11171962	1701	
637.5	6	BYLER	BRANDT	E	11171962	1701	122919
637.1	6	GUEST	ALLEYN	WILLIAM	12271969	2701	
637.1	6	GUEST	ALLEYN	W	12271969	2701	062819
636.6	6	WAYERSKI	TORY	KENDAL	12111962	3702	
636.6	6	WAYERSKI	TORY	K	12111962	3702	080619
636.5	6	BAUERMEISTER	CHET	D	07101963	1100	
636.5	6	BAUERMEISTER	CHET	D	07101963	1100	010419
636.5	6	LOUANGAPHAY	KHONESAVANH		12081962	2701	
636.5	6	LOUANGAPHAY	KHONESAVANH		12081962	2701	080419

\* Screenshot contains only public, disclosable data.

# ModifyLinkPlus.sas Can Itself be Modified

- **The hardcoded 15 Link Plus classes can be redefined.**
- **New classes can be created. For example:**
  - **A class just for those potential matched pairs of records with a January 1<sup>st</sup> birthday.**
  - **One or more classes to help discriminate and isolate potential twins with similar first names.**

# Processing the Next Batch *Faster*

- **After linking a batch of data once, can determine cutoff values by class.**
- **Can use these cutoff values with `SeparateByCondition.sas` to programmatically**
  - 1. accept “good” matched pairs,**
  - 2. reject “bad” matched pairs,**
  - 3. create a Link Plus report file that contains only uncertain matches.**

# Determination of Cutoff Values by Class

<b>Class</b>	<b>Total</b>	<b>Accepted</b>	<b>Rejected</b>	<b>Uncertain</b>	<b>Min Score Accepted</b>	<b>Max Score Not Accepted</b>
6	2,267	2,264	3	0	13.3	16.3
7	417	124	292	1	10.2	33.6
8	3,328	82	3,246	0	13.6	20.6
9	25,746	24	25,722	0	12.0	28.2
10	54,875	4	54,871	0	22.4	24.1

# Summary of Link Plus Toolkit Benefits

- **By modifying scores, can accept and reject matches by score *and* class.**
- **Can redefine the 15 default Link Plus classes.**
- **Can create new classes.**
- **Through the creation of new classes, can tease out islands of good or bad potential matches from other potential matches.**
- **Once cutoff scores are established for a linkage project, the need for manual review is greatly reduced when linking additional data.**

# How Does it End?

## Washington State Top 5 Leading Causes of Death, 2001-2010\* Divorced Men Born Between 1962 and 1971

ICD-10	Count	Description
X44	143	Accidental poisoning by and exposure to other and unspecified drugs, medicaments and biological substances
X42	118	Accidental poisoning by and exposure to narcotics and psychodysleptics [hallucinogens], not elsewhere classified
X72	88	Intentional self-harm by handgun discharge
I25	86	Chronic ischemic heart disease
K70	84	Alcoholic liver disease

\* Draft counts

# CONTACT ME

John Sabel

[john.sabel@ofm.wa.gov](mailto:john.sabel@ofm.wa.gov)

Washington Education Research & Data Center

210 11th Ave SW, Room 318

P.O. Box 43113

Olympia, WA 98504-3113

Phone: (360) 902-0599

ERDC website: [www.erdc.wa.gov](http://www.erdc.wa.gov)

## Program Flow for the Initial Linkage of Death and Divorce Data

```
%let L_Drive=L:\;

/* Read in Link Plus Linkage Report into a SAS data set. */
%ReadInLinkPlusReport(LinkRptPathIn=&L_Drive.\Report\Unmodified\DeathToDivorce.txt,
                      MetadataDsnOut=DeathToDivorce_MetaData,
                      ReportDsnOut=DeathToDivorce,
                      Debug=N)

/*
Reassign score so that each class has a non-overlapping set of scores.
*/
%ModifyLinkPlusReport(ReportDsnIn=DeathToDivorce,
                     ReportDsnOut=DeathToDivorce_Modified,
                     Debug=N)

/*
Get a feel for the data by determining the number of potential matches by class.
*/
proc freq data=DeathToDivorce_Modified;
    title "DeathToDivorce_Modified";
    table class / nopercnt;
run;

/* Create a modified Link Plus Linkage Report with the modified scores. */
%AssembleLinkPlusReport(MetadataDsnIn=DeathToDivorce_MetaData,
                       ReportDsnIn=DeathToDivorce_Modified,

                       LinkRptPathOut=&L_Drive.\Report\DeathToDivorce_Modified.txt,
                       Debug=N)

/*****
*
* Now perform manual review on the modified Link Plus Linkage report.
*
*****/

/*
Merge all the Link Plus tab-delimited export files into the final single SAS
data set.
*/
%MergeLinkPlusLinkageFiles(DirIn=&L_Drive.\Export,
                          DsnOut=DeathToDivorce_Linkages,
                          OnlyGoodLinks=N,
                          Debug=N)

/*
Post-linkage analysis:
    1) Counts of matches accepted, rejected, uncertain.
    2) Minimum score by class accepted. Below these scores ALL matches
       were rejected.
    3) Maximum score by class rejected. Above these scores ALL matches
       were accepted.
```

The cutoff values by class can be used with to streamline future data linkages that involve data sets with similar characteristics in their linking variables.

```
*/  
proc sql;  
    title 'DeathToDivorce_Linkages Cutoff Scores by Class';  
    SELECT class,  
        count(*) AS TotalMatches,  
        sum(case when Match_Status = 1 then 1 else 0 end) as AcceptedMatches,  
        sum(case when Match_Status = 2 then 1 else 0 end) as RejectedMatches,  
        sum(case when Match_Status = 3 then 1 else 0 end) as UncertainMatches,  
        min(case when Match_Status = 1 then  
(score - (class * 100)) end) AS MinScoreAccepted,  
        max(case when Match_Status <> 1 then  
(score - (class * 100)) end) AS MaxScoreNotAccepted  
    FROM DeathToDivorce_Linkages  
    GROUP BY class  
    ORDER BY class  
    ;  
quit;
```

## Program Flow for a Follow Up Linkage of Additional Death and Divorce Data

```
%let L_Drive=L:\;

/* Read in Link Plus Linkage Report into a SAS data set. */
%ReadInLinkPlusReport(LinkRptPathIn=&L_Drive.\Report\Unmodified\DeathToDivorce.txt,
                      MetadataDsnOut=DeathToDivorce_MetaData,
                      ReportDsnOut=DeathToDivorce,
                      Debug=N)

/*
Reassign score so that each class has a non-overlapping set of scores.
*/
%ModifyLinkPlusReport(ReportDsnIn=DeathToDivorce,
                     ReportDsnOut=DeathToDivorce_Modified,
                     Debug=N)

/*
Get a feel for the data by determining the number of potential matches by class.
*/
proc freq data=DeathToDivorce_Modified;
    title "DeathToDivorce_Modified";
    table class / nopercnt;
run;

/*
Separate modified linkage report into three data sets:
1) TheGood: Potential matches accepted sight unseen.
2) TheBad: Potential matches rejected sight unseen.
3) TheUgly: Potential matches that will require manual review.
*/

%SeparateByCondition(In=DeathToDivorce_Modified,
MeetsConditionOut=TheGood,
DoesNotMeetConditionOut=NotSoGood,
Condition=
(class = 6 AND orig_score >= 16.5) OR
(class = 7 AND orig_score >= 33.9) OR
(class = 8 AND orig_score >= 20.9) OR
(class = 9 AND orig_score >= 28.5) OR
(class = 10 AND orig_score >= 24.4))

%SeparateByCondition(In=NotSoGood,
MeetsConditionOut=TheBad,
DoesNotMeetConditionOut=TheUgly,
Condition=
(class = 7 AND orig_score <= 10.0) OR
(class = 8 AND orig_score <= 13.3) OR
(class = 9 AND orig_score <= 11.7) OR
(class = 10 AND orig_score <= 22.1))

/*
Create export files for the data sets "TheGood" and "TheBad", setting appropriate
Match_Status values for each.
*/

%CreateLinkPlusExportFile(ReportDsnIn=TheGood,
FilePathOut=&L_Drive.\Report\TheGood.txt,
Match_Status=1,
Key_1=certno,
```

```

Key_2=certnum,
Debug=N)

%CreateLinkPlusExportFile(ReportDsnIn=TheBad,
FilePathOut=&L_Drive.\Report\TheBad.txt,
Match_Status=2,
Key_1=certno,
Key_2=certnum,
Debug=N)

/*
Create a modified Link Plus Linkage Report with the modified scores for only the
potential match pairs that require manual review, TheUgly.
*/
%AssembleLinkPlusReport(MetadataDsnIn=DeathToDivorce_MetaData,
                        ReportDsnIn=TheUgly,
                        LinkRptPathOut=&L_Drive.\Report\TheUgly.txt,
                        Debug=N)

/*****
*
* Now perform manual review on the modified Link Plus Linkage report "The
* Ugly" and create the appropriate export files.
*
*****/

/*
Merge all the Link Plus tab-delimited export files into the final single SAS
data set. These files will include TheGood.txt, TheBad.txt, and the export files
from manual.
*/
%MergeLinkPlusLinkageFiles(DirIn=&L_Drive.\Export,
                           DsnOut=DeathToDivorce_Linkages,
                           OnlyGoodLinks=N,
                           Debug=N)

/*
Post-linkage analysis:
    1) Counts of matches accepted, rejected, uncertain.
    2) Minimum score by class accepted. Below these scores ALL matches
       were rejected.
    3) Maximum score by class rejected. Above these scores ALL matches
       were accepted.
*/
proc sql;
    title 'DeathToDivorce_Linkages Cutoff Scores by Class';
    SELECT class,
           count(*) AS TotalMatches,
           sum(case when Match_Status = 1 then 1 else 0 end) as AcceptedMatches,
           sum(case when Match_Status = 2 then 1 else 0 end) as RejectedMatches,
           sum(case when Match_Status = 3 then 1 else 0 end) as UncertainMatches,
           min(case when Match_Status = 1 then
(score - (class * 100)) end) AS MinScoreAccepted,
           max(case when Match_Status <> 1 then
(score - (class * 100)) end) AS MaxScoreNotAccepted
    FROM DeathToDivorce_Linkages
    GROUP BY class
    ORDER BY class
    ;
quit;

```

## Link Plus Tips

### Map the root directory of every Link Plus project to a single drive letter

1. Consistently mapping to a single drive letter makes Link Plus easier to use for two reasons:
  - a. It allows you to more easily navigate to your project's Link Plus folders, (e.g., Configuration, Export, Report).
    - i. Link Plus has its own set of default paths to these folders that are based off the C drive. It really likes these paths. If you don't use Link Plus's default in the install directory, you will forever be copying directory paths of the actual location of these folders from Windows Explorer to Link Plus.
  - b. It makes Link Plus projects portable.
    - i. Link Plus stores directory paths in a project's configuration file and in the Linkage Report. So if a project and its associated folders are moved to another location, Link Plus will no longer work with the project. If you had to, you could open the configuration file and the Linkage Report file in a text editor and modify the pertinent paths, but this can be error prone. Mapping to the same drive letter from the beginning makes using Link Plus more pleasant in the long run.
2. How to setup Link Plus and map projects to a single drive letter:
  - a. Choose a drive letter, like "L," to associate with any and all Link Plus projects. Then place shortcuts in each default Link Plus directory on your C drive to the analogous directory immediately off the chosen drive letter.
    - i. For example, if your chosen drive letter is "L," in the **C: \RegPlus\LinkPlus\Export** folder put a shortcut to **L: \Export**.
  - b. Then before beginning any Link Plus project, map the location of the root directory of the project to that drive letter.
    - i. For location on network drives, you can use Windows Explorer to map network drives.
      1. For example, if you wanted to start a new Link Plus project in the **//DohServer/Office/Section/LinkPlus**, then use Windows Explorer to map the L drive to this location.
    - ii. For locations on a local hard drive, you have to use the DOS command **subst**.
      1. For example, to assign the directory **C: \LinkPlus** to the L drive, in a command window execute the following command:

```
subst L: C: \LinkPlus
```

### Clean all linking fields in a consistent manner

1. With new data, work with it to determine to come up with a set of business rules on how it should be cleaned. These business rules could be such things as make all characters upper case,

remove all dashes, and so on. As you code the business rules, and run the data though, more ways of cleaning the data will likely present themselves. Especially with name data, this could be an iterative process.

2. One implementation of a SAS macro to clean name data is the **%StandardizeNames** macro. It can be found on the ERDC website at [www.erd.c.wa.gov/briefs/technical](http://www.erd.c.wa.gov/briefs/technical).

### **When linking two files, and only one is deduplicated down to a single record per person, make that one “File 1”**

1. When linking two files and only one file is rigorously deduplicated, make the deduplicated file “File 1” in the configuration file. Then each deduplicated record in “File 1” will match 0 to many records in “File 2.”
2. If the deduplicated file is made “File 2,” then each deduplicated record should match 0 to 1 records in “File 1.”

### **Use the ERDC Link Plus Toolkit set of SAS macros to increase the power and flexibility of Link Plus**

1. One of Link Plus’s limitations is its inflexibility in assigning the “Class” variable. The “Class” variable is used in manual review. It is based on the concordance between the following variables: first name, last name, birth date, and social security number. In the process of running Link Plus, each possible matched pair is assigned a “Class” value, which ranges from 1 to 15. Link Plus does not allow you to redefine “Class” variable definitions, nor does it allow you to define additional classes. But the macros **%ReadInLinkPlus**, **%ModifyLinkPlusReport** and **%AssembleLinkPlus** do allow you to redefine “Class” variable definitions, as well as define new classes. In addition, these macros allow you to modify the “Score” variable.
2. These macros are available bundled together as **LinkPlusToolkit.sas** on the ERDC website at [www.erd.c.wa.gov/briefs/technical](http://www.erd.c.wa.gov/briefs/technical).

### **An excellent, though unofficial Link Plus manual**

1. [www.npaihb.org/images/epicenter\\_docs/NW-Idea/Link Plus manual 2011\\_FINAL.docx](http://www.npaihb.org/images/epicenter_docs/NW-Idea/Link%20Plus%20manual%202011_FINAL.docx)

### **Comments?**

1. For comments, corrections, or more information, contact John Sabel, 360-902-0943, [john.sabel@ofm.wa.gov](mailto:john.sabel@ofm.wa.gov)