

Version

3.0

pdGender User Guide

Name and Gender Coding Database

A one-of-a-kind proprietary resource designed to facilitate gender coding name lists. The package also includes extensive language of origin and use information, and the *Pro* edition includes fuzzy logic to match typographical errors. Ancestry researchers, students, teachers, and scholars benefit as well because this software is recommended for study in genealogy, onomatology, anthroponymy, ethnology, linguistics, and related disciplines.



Peacock Data, Inc.

California, USA ☎ 800-609-9231

Web address: www.peacockdata.com



TABLE OF CONTENTS

Introduction.....	4
Importing Data Into Your System	5
Included Database Files	5
File Formats	5
Character Set	6
File Layouts and Data Definitions	7
Layout of pdGender Main Gender Database	7
Layout of pdGender Fuzzy Logic Gender Database (PRO ONLY).....	12
Using the Main Gender Database.....	16
PEACOCK_ID Field	16
Name Field	16
Gender	19
GIVEN and NICK Fields	19
Name Rank.....	20
Archaic Names	21
Language.....	21
Special and Unique Origins	26
WORLD Field	28
Filtered Gender Coding Fields.....	29
Fuzzy Logic (Pro only)	32
Using the Fuzzy Logic File.....	32
Compatibility	35
Using pdGender with pdNickname and pdSurname.....	35
User Guide Updates.....	35

Database Version Number.....35

Site License36

Copyright Notice.....36

INTRODUCTION



Male and female identification is essential for businesses and organizations. It allows you to send mail with a personal touch. Gender Coding also allows you to filter, map, and analyze your data based on this critical demographic.

pdGender lets you accomplish this in ways not before possible on this scale.

Coverage includes hundreds of thousands of names and the package employs the best matching algorithms designed for this process. The software is a one-of-a-kind proprietary resource that for more than 20 years has been utilized by businesses and organizations around the world in applications you use every day.

But what makes this gender coding database truly different are 140 special gender coding fields filtered for languages, rare usage by one gender, archaic names, and nicknames. With this software:

- When names are different genders in different languages and nationalities, users can choose which languages and nationalities to take precedence
- When names are one gender now in current times, but were the opposite gender in a previous era, the system automatically applies the modern usage
- When names are one gender when used as a proper given name and unisex when used as a nickname, users can choose to have the given name usage applied
- When unisex names are only rarely used by one gender and much more common in the opposite gender, users can choose to ignore the rare instances

As an added benefit, languages of origin and use have also been researched and included, and the enhanced version even incorporates sophisticated fuzzy logic which allows matching when lists have typographical errors.

This easy-to-use, comprehensive, and up-to-date software is of great value for businesses and organizations working with lists of names, but ancestry researchers, students, teachers, and scholars benefit as well because this software is recommended for study in genealogy, onomatology, anthroponymy, ethnology, linguistics, and related disciplines.

pdGender is available in **Pro** and **Standard** editions. This guide covers both versions.

✓ **PRO:** software includes more than 397,000 regular gender coding records with languages of origin and use and more than 3 million fuzzy logic records. The file includes one unfiltered and 140 filtered gender coding fields.

✓ **STANDARD:** software includes more than 397,000 regular gender coding records with languages of origin and use. The file includes one unfiltered and 140 filtered gender coding fields.

IMPORTING DATA INTO YOUR SYSTEM

pdGender is designed to be compatible with any database system. It comes in multiple file formats, uses only the ANSI character set, and has a well-defined layout.

INCLUDED DATABASE FILES

pdGender has one or two data sets depending on the version. There is a main gender coding database and, in the *Pro* edition only, a fuzzy logic gender coding file.

Included files are:

MAIN GENDER FILE

This data set lists more than 397,000 given name and nickname formations along with one unfiltered and 140 filtered gender coding fields. It also includes a name ranking, archaic flag, and languages of origin and use.

FUZZY LOGIC GENDER FILE (PRO ONLY)

This data set lists more than 3 million fuzzy logic name formations along with one unfiltered and 140 filtered gender coding fields. It is available in the *Pro* edition only.

FILE FORMATS

The database is available in three common file formats. Each format contains the same data.

Available file formats are:

CSV-COMMA SEPARATED VALUES

Files in Comma Separated Values (CSV) format (also known as Comma Delimited) separate fields with commas, and alpha/numeric character fields are usually delimited with double quotes (in case some of the field content includes commas). This format is the most commonly used. It is a native format for Microsoft Excel and is compatible with nearly all database management systems and spreadsheets.

TXT-FIXED LENGTH

Files in Fixed Length (TXT) format (also known as Standard Data Format or SDF) use constant field positions and lengths for all records. In other words, each field starts and ends at the same place in the text file and each record is on a separate line. While not as popular as comma separated values, this format is preferred by many due to its input precision and is widely used to transfer data between different software programs. It is compatible with most database management systems and spreadsheets.

DBF-DATABASE

Files in DBF database format (also known as xBase) are native to Microsoft FoxPro and Visual FoxPro, dataBased Intelligence dBase, Alaska Software XBase++, Apollo Database Engine, Apycom Software DBFView, Astersoft DBF Manager, DS-Datasoft Visual DBU, Elsoft DBF Commander, GrafX Software Clipper and Vulcan.NET, Multisoft FlagShip, Recital Software Recital, Software Perspectives Cule.Net, and xHarbour.com xHarbour. They are also compatible with any database management system that can import the DBF (xBase) format, such as Microsoft Access, Microsoft SQL Server, and numerous others.

CHARACTER SET

The ANSI character set is utilized for all database records. This includes ASCII values 0 to 127 and extended values 128 to 255. These are also known as the extended Latin alphabet. Some users may need to configure their database system to import the extended values. In many cases the option will be labeled the “Latin-1” character set.

FILE LAYOUTS AND DATA DEFINITIONS

Below are the complete layout specifications and data definitions of all files provided with *pdGender*.

Each line below contains the following information: **FIELD NUMBER**: field position number. **FIELD NAME**: name of field. **FIELD TYPE**: field data type; “Chr” = alpha/numeric characters, “Num” = numbers. **FIELD LENGTH**: length of field. **DECIMAL PLACES**: number of decimal places (if any). **START POSITION**: field starting position. **END POSITION**: field ending position. **DESCRIPTION**: data definition of field contents.

LAYOUT OF PDGENDER MAIN GENDER DATABASE

Field Count: 150

Total Length: 556

Record Count: Pro and Standard: 397,847

FIELD NUMBER	FIELD NAME	FIELD TYPE	FIELD LENGTH	DECIMAL PLACES	START POSITION	END POSITION	DESCRIPTION
1	PEACOCK_ID	Chr	17		1	17	Unique identifier for each record
2	NAME	Chr	35		18	52	Stylized name spelling
3	GENDER	Chr	1		53	53	Gender flag: <i>M = Male</i> <i>F = Female</i>
4	GIVEN	Chr	1		54	54	Given name flag: G = Is a given name
5	NICK	Chr	1		55	55	Nickname flag: N = Is a nickname
6	RANK	Num	5	0	56	60	Name rank in the United States
7	ARCHAIC	Chr	1		61	61	Archaic name flag: <i>A = Archaic</i>
8	LANGUAGE	Chr	254		62	315	Language or languages of origin and use
9	SPECIAL	Chr	100		316	415	Special and unique origins
10	WORLD	Chr	1		416	416	Unfiltered gender flag: <i>M = Male</i> <i>F = Female</i> <i>U = Unisex</i>
11	WORLD_XA	Chr	1		417	417	Filtered gender #1: <i>M = Male</i> <i>F = Female</i> <i>U = Unisex</i> <i>NOTE: All remaining filtered gender fields also use this key.</i>
12	WORLD_XAN	Chr	1		418	418	Filtered gender #2
13	WORLD_XAR	Chr	1		419	419	Filtered gender #3
14	WORLD_XANR	Chr	1		420	420	Filtered gender #4
15	USA_XA	Chr	1		421	421	Filtered gender #5
16	USA_XAN	Chr	1		422	422	Filtered gender #6
17	USA_XAR	Chr	1		423	423	Filtered gender #7

18	USA_XANR	Chr	1		424	424	Filtered gender #8
19	US_ES_XA	Chr	1		425	425	Filtered gender #9
20	US_ES_XAN	Chr	1		426	426	Filtered gender #10
21	US_ES_XAR	Chr	1		427	427	Filtered gender #11
22	US_ES_XANR	Chr	1		428	428	Filtered gender #12
23	US_HS_XA	Chr	1		429	429	Filtered gender #13
24	US_HS_XAN	Chr	1		430	430	Filtered gender #14
25	US_HS_XAR	Chr	1		431	431	Filtered gender #15
26	US_HS_XANR	Chr	1		432	432	Filtered gender #16
27	US_FR_XA	Chr	1		433	433	Filtered gender #17
28	US_FR_XAN	Chr	1		434	434	Filtered gender #18
29	US_FR_XAR	Chr	1		435	435	Filtered gender #19
30	US_FR_XANR	Chr	1		436	436	Filtered gender #20
31	ENG_XA	Chr	1		437	437	Filtered gender #21
32	ENG_XAN	Chr	1		438	438	Filtered gender #22
33	ENG_XAR	Chr	1		439	439	Filtered gender #23
34	ENG_XANR	Chr	1		440	440	Filtered gender #24
35	EN_AA_XA	Chr	1		441	441	Filtered gender #25
36	EN_AA_XAN	Chr	1		442	442	Filtered gender #26
37	EN_AA_XAR	Chr	1		443	443	Filtered gender #27
38	EN_AA_XANR	Chr	1		444	444	Filtered gender #28
39	EN_ES_XA	Chr	1		445	445	Filtered gender #29
40	EN_ES_XAN	Chr	1		446	446	Filtered gender #30
41	EN_ES_XAR	Chr	1		447	447	Filtered gender #31
42	EN_ES_XANR	Chr	1		448	448	Filtered gender #32
43	EN_HS_XA	Chr	1		449	449	Filtered gender #33
44	EN_HS_XAN	Chr	1		450	450	Filtered gender #34
45	EN_HS_XAR	Chr	1		451	451	Filtered gender #35
46	EN_HS_XANR	Chr	1		452	452	Filtered gender #36
47	EN_FR_XA	Chr	1		453	453	Filtered gender #37
48	EN_FR_XAN	Chr	1		454	454	Filtered gender #38
49	EN_FR_XAR	Chr	1		455	455	Filtered gender #39
50	EN_FR_XANR	Chr	1		456	456	Filtered gender #40
51	AFRAM_XA	Chr	1		457	457	Filtered gender #41
52	AFRAM_XAN	Chr	1		458	458	Filtered gender #42
53	AFRAM_XAR	Chr	1		459	459	Filtered gender #43
54	AFRAM_XANR	Chr	1		460	460	Filtered gender #44
55	SPA_XA	Chr	1		461	461	Filtered gender #45
56	SPA_XAN	Chr	1		462	462	Filtered gender #46

57	SPA_XAR	Chr	1		463	463	Filtered gender #47
58	SPA_XANR	Chr	1		464	464	Filtered gender #48
59	HISP_XA	Chr	1		465	465	Filtered gender #49
60	HISP_XAN	Chr	1		466	466	Filtered gender #50
61	HISP_XAR	Chr	1		467	467	Filtered gender #51
62	HISP_XANR	Chr	1		468	468	Filtered gender #52
63	FRA_XA	Chr	1		469	469	Filtered gender #53
64	FRA_XAN	Chr	1		470	470	Filtered gender #54
65	FRA_XAR	Chr	1		471	471	Filtered gender #55
66	FRA_XANR	Chr	1		472	472	Filtered gender #56
67	AFR_XA	Chr	1		473	473	Filtered gender #57
68	AFR_XAN	Chr	1		474	474	Filtered gender #58
69	AFR_XAR	Chr	1		475	475	Filtered gender #59
70	AFR_XANR	Chr	1		476	476	Filtered gender #60
71	BRIT_XA	Chr	1		477	477	Filtered gender #61
72	BRIT_XAN	Chr	1		478	478	Filtered gender #62
73	BRIT_XAR	Chr	1		479	479	Filtered gender #63
74	BRIT_XANR	Chr	1		480	480	Filtered gender #64
75	CEL_XA	Chr	1		481	481	Filtered gender #65
76	CEL_XAN	Chr	1		482	482	Filtered gender #66
77	CEL_XAR	Chr	1		483	483	Filtered gender #67
78	CEL_XANR	Chr	1		484	484	Filtered gender #68
79	EASIA_XA	Chr	1		485	485	Filtered gender #69
80	EASIA_XAN	Chr	1		486	486	Filtered gender #70
81	EASIA_XAR	Chr	1		487	487	Filtered gender #71
82	EASIA_XANR	Chr	1		488	488	Filtered gender #72
83	EA_PI_XA	Chr	1		489	489	Filtered gender #73
84	EA_PI_XAN	Chr	1		490	490	Filtered gender #74
85	EA_PI_XAR	Chr	1		491	491	Filtered gender #75
86	EA_PI_XANR	Chr	1		492	492	Filtered gender #76
87	GAEL_XA	Chr	1		493	493	Filtered gender #77
88	GAEL_XAN	Chr	1		494	494	Filtered gender #78
89	GAEL_XAR	Chr	1		495	495	Filtered gender #79
90	GAEL_XANR	Chr	1		496	496	Filtered gender #80
91	DEU_XA	Chr	1		497	497	Filtered gender #81
92	DEU_XAN	Chr	1		498	498	Filtered gender #82
93	DEU_XAR	Chr	1		499	499	Filtered gender #83
94	DEU_XANR	Chr	1		500	500	Filtered gender #84
95	GEM_XA	Chr	1		501	501	Filtered gender #85

96	GEM_XAN	Chr	1		502	502	Filtered gender #86
97	GEM_XAR	Chr	1		503	503	Filtered gender #87
98	GEM_XANR	Chr	1		504	504	Filtered gender #88
99	HAW_XA	Chr	1		505	505	Filtered gender #89
100	HAW_XAN	Chr	1		506	506	Filtered gender #90
101	HAW_XAR	Chr	1		507	507	Filtered gender #91
102	HAW_XANR	Chr	1		508	508	Filtered gender #92
103	IND_XA	Chr	1		509	509	Filtered gender #93
104	IND_XAN	Chr	1		510	510	Filtered gender #94
105	IND_XAR	Chr	1		511	511	Filtered gender #95
106	IND_XANR	Chr	1		512	512	Filtered gender #96
107	ITA_XA	Chr	1		513	513	Filtered gender #97
108	ITA_XAN	Chr	1		514	514	Filtered gender #98
109	ITA_XAR	Chr	1		515	515	Filtered gender #99
110	ITA_XANR	Chr	1		516	516	Filtered gender #100
111	JW_XA	Chr	1		517	517	Filtered gender #101
112	JW_XAN	Chr	1		518	518	Filtered gender #102
113	JW_XAR	Chr	1		519	519	Filtered gender #103
114	JW_XANR	Chr	1		520	520	Filtered gender #104
115	MUS_XA	Chr	1		521	521	Filtered gender #105
116	MUS_XAN	Chr	1		522	522	Filtered gender #106
117	MUS_XAR	Chr	1		523	523	Filtered gender #107
118	MUS_XANR	Chr	1		524	524	Filtered gender #108
119	NATAM_XA	Chr	1		525	525	Filtered gender #109
120	NATAM_XAN	Chr	1		526	526	Filtered gender #110
121	NATAM_XAR	Chr	1		527	527	Filtered gender #111
122	NATAM_XANR	Chr	1		528	528	Filtered gender #112
123	PISLR_XA	Chr	1		529	529	Filtered gender #113
124	PISLR_XAN	Chr	1		530	530	Filtered gender #114
125	PISLR_XAR	Chr	1		531	531	Filtered gender #115
126	PISLR_XANR	Chr	1		532	532	Filtered gender #116
127	ROA_XA	Chr	1		533	533	Filtered gender #117
128	ROA_XAN	Chr	1		534	534	Filtered gender #118
129	ROA_XAR	Chr	1		535	535	Filtered gender #119
130	ROA_XANR	Chr	1		536	536	Filtered gender #120
131	SCAND_XA	Chr	1		537	537	Filtered gender #121
132	SCAND_XAN	Chr	1		538	538	Filtered gender #122
133	SCAND_XAR	Chr	1		539	539	Filtered gender #123
134	SCAND_XANR	Chr	1		540	540	Filtered gender #124

135	SLA_XA	Chr	1		541	541	Filtered gender #125
136	SLA_XAN	Chr	1		542	542	Filtered gender #126
137	SLA_XAR	Chr	1		543	543	Filtered gender #127
138	SLA_XANR	Chr	1		544	544	Filtered gender #128
139	CYM_XA	Chr	1		545	545	Filtered gender #129
140	CYM_XAN	Chr	1		546	546	Filtered gender #130
141	CYM_XAR	Chr	1		547	547	Filtered gender #131
142	CYM_XANR	Chr	1		548	548	Filtered gender #132
143	WEST_XA	Chr	1		549	549	Filtered gender #133
144	WEST_XAN	Chr	1		550	550	Filtered gender #134
145	WEST_XAR	Chr	1		551	551	Filtered gender #135
146	WEST_XANR	Chr	1		552	552	Filtered gender #136
147	NWEST_XA	Chr	1		553	553	Filtered gender #137
148	NWEST_XAN	Chr	1		554	554	Filtered gender #138
149	NWEST_XAR	Chr	1		555	555	Filtered gender #139
150	NWEST_XANR	Chr	1		556	556	Filtered gender #140

LAYOUT OF PDGENDER FUZZY LOGIC GENDER DATABASE (PRO ONLY)

Field Count: 142

Total Length: 176

Record Count: 3,171,385

FIELD NUMBER	FIELD NAME	FIELD TYPE	FIELD LENGTH	DECIMAL PLACES	START POSITION	END POSITION	DESCRIPTION
1	NAME	Chr	35		1	35	Fuzzy logic name spelling
2	WORLD	Chr	1		36	36	Unfiltered gender flag: <i>M = Male</i> <i>F = Female</i> <i>U = Unisex</i>
3	WORLD_XA	Chr	1		37	37	Filtered gender #1: <i>M = Male</i> <i>F = Female</i> <i>U = Unisex</i> <i>NOTE: All remaining filtered gender fields also use this key.</i>
4	WORLD_XAN	Chr	1		38	38	Filtered gender #2
5	WORLD_XAR	Chr	1		39	39	Filtered gender #3
6	WORLD_XANR	Chr	1		40	40	Filtered gender #4
7	USA_XA	Chr	1		41	41	Filtered gender #5
8	USA_XAN	Chr	1		42	42	Filtered gender #6
9	USA_XAR	Chr	1		43	43	Filtered gender #7
10	USA_XANR	Chr	1		44	44	Filtered gender #8
11	US_ES_XA	Chr	1		45	45	Filtered gender #9
12	US_ES_XAN	Chr	1		46	46	Filtered gender #10
13	US_ES_XAR	Chr	1		47	47	Filtered gender #11
14	US_ES_XANR	Chr	1		48	48	Filtered gender #12
15	US_HS_XA	Chr	1		49	49	Filtered gender #13
16	US_HS_XAN	Chr	1		50	50	Filtered gender #14
17	US_HS_XAR	Chr	1		51	51	Filtered gender #15
18	US_HS_XANR	Chr	1		52	52	Filtered gender #16
19	US_FR_XA	Chr	1		53	53	Filtered gender #17
20	US_FR_XAN	Chr	1		54	54	Filtered gender #18
21	US_FR_XAR	Chr	1		55	55	Filtered gender #19
22	US_FR_XANR	Chr	1		56	56	Filtered gender #20
23	ENG_XA	Chr	1		57	57	Filtered gender #21
24	ENG_XAN	Chr	1		58	58	Filtered gender #22
25	ENG_XAR	Chr	1		59	59	Filtered gender #23
26	ENG_XANR	Chr	1		60	60	Filtered gender #24
27	EN_AA_XA	Chr	1		61	61	Filtered gender #25

28	EN_AA_XAN	Chr	1		62	62	Filtered gender #26
29	EN_AA_XAR	Chr	1		63	63	Filtered gender #27
30	EN_AA_XANR	Chr	1		64	64	Filtered gender #28
31	EN_ES_XA	Chr	1		65	65	Filtered gender #29
32	EN_ES_XAN	Chr	1		66	66	Filtered gender #30
33	EN_ES_XAR	Chr	1		67	67	Filtered gender #31
34	EN_ES_XANR	Chr	1		68	68	Filtered gender #32
35	EN_HS_XA	Chr	1		69	69	Filtered gender #33
36	EN_HS_XAN	Chr	1		70	70	Filtered gender #34
37	EN_HS_XAR	Chr	1		71	71	Filtered gender #35
38	EN_HS_XANR	Chr	1		72	72	Filtered gender #36
39	EN_FR_XA	Chr	1		73	73	Filtered gender #37
40	EN_FR_XAN	Chr	1		74	74	Filtered gender #38
41	EN_FR_XAR	Chr	1		75	75	Filtered gender #39
42	EN_FR_XANR	Chr	1		76	76	Filtered gender #40
43	AFRAM_XA	Chr	1		77	77	Filtered gender #41
44	AFRAM_XAN	Chr	1		78	78	Filtered gender #42
45	AFRAM_XAR	Chr	1		79	79	Filtered gender #43
46	AFRAM_XANR	Chr	1		80	80	Filtered gender #44
47	SPA_XA	Chr	1		81	81	Filtered gender #45
48	SPA_XAN	Chr	1		82	82	Filtered gender #46
49	SPA_XAR	Chr	1		83	83	Filtered gender #47
50	SPA_XANR	Chr	1		84	84	Filtered gender #48
51	HISP_XA	Chr	1		85	85	Filtered gender #49
52	HISP_XAN	Chr	1		86	86	Filtered gender #50
53	HISP_XAR	Chr	1		87	87	Filtered gender #51
54	HISP_XANR	Chr	1		88	88	Filtered gender #52
55	FRA_XA	Chr	1		89	89	Filtered gender #53
56	FRA_XAN	Chr	1		90	90	Filtered gender #54
57	FRA_XAR	Chr	1		91	91	Filtered gender #55
58	FRA_XANR	Chr	1		92	92	Filtered gender #56
59	AFR_XA	Chr	1		93	93	Filtered gender #57
60	AFR_XAN	Chr	1		94	94	Filtered gender #58
61	AFR_XAR	Chr	1		95	95	Filtered gender #59
62	AFR_XANR	Chr	1		96	96	Filtered gender #60
63	BRIT_XA	Chr	1		97	97	Filtered gender #61
64	BRIT_XAN	Chr	1		98	98	Filtered gender #62
65	BRIT_XAR	Chr	1		99	99	Filtered gender #63
66	BRIT_XANR	Chr	1		100	100	Filtered gender #64

67	CEL_XA	Chr	1		101	101	Filtered gender #65
68	CEL_XAN	Chr	1		102	102	Filtered gender #66
69	CEL_XAR	Chr	1		103	103	Filtered gender #67
70	CEL_XANR	Chr	1		104	104	Filtered gender #68
71	EASIA_XA	Chr	1		105	105	Filtered gender #69
72	EASIA_XAN	Chr	1		106	106	Filtered gender #70
73	EASIA_XAR	Chr	1		107	107	Filtered gender #71
74	EASIA_XANR	Chr	1		108	108	Filtered gender #72
75	EA_PI_XA	Chr	1		109	109	Filtered gender #73
76	EA_PI_XAN	Chr	1		110	110	Filtered gender #74
77	EA_PI_XAR	Chr	1		111	111	Filtered gender #75
78	EA_PI_XANR	Chr	1		112	112	Filtered gender #76
79	GAEL_XA	Chr	1		113	113	Filtered gender #77
80	GAEL_XAN	Chr	1		114	114	Filtered gender #78
81	GAEL_XAR	Chr	1		115	115	Filtered gender #79
82	GAEL_XANR	Chr	1		116	116	Filtered gender #80
83	DEU_XA	Chr	1		117	117	Filtered gender #81
84	DEU_XAN	Chr	1		118	118	Filtered gender #82
85	DEU_XAR	Chr	1		119	119	Filtered gender #83
86	DEU_XANR	Chr	1		120	120	Filtered gender #84
87	GEM_XA	Chr	1		121	121	Filtered gender #85
88	GEM_XAN	Chr	1		122	122	Filtered gender #86
89	GEM_XAR	Chr	1		123	123	Filtered gender #87
90	GEM_XANR	Chr	1		124	124	Filtered gender #88
91	HAW_XA	Chr	1		125	125	Filtered gender #89
92	HAW_XAN	Chr	1		126	126	Filtered gender #90
93	HAW_XAR	Chr	1		127	127	Filtered gender #91
94	HAW_XANR	Chr	1		128	128	Filtered gender #92
95	IND_XA	Chr	1		129	129	Filtered gender #93
96	IND_XAN	Chr	1		130	130	Filtered gender #94
97	IND_XAR	Chr	1		131	131	Filtered gender #95
98	IND_XANR	Chr	1		132	132	Filtered gender #96
99	ITA_XA	Chr	1		133	133	Filtered gender #97
100	ITA_XAN	Chr	1		134	134	Filtered gender #98
101	ITA_XAR	Chr	1		135	135	Filtered gender #99
102	ITA_XANR	Chr	1		136	136	Filtered gender #100
103	JW_XA	Chr	1		137	137	Filtered gender #101
104	JW_XAN	Chr	1		138	138	Filtered gender #102
105	JW_XAR	Chr	1		139	139	Filtered gender #103

106	JW_XANR	Chr	1		140	140	Filtered gender #104
107	MUS_XA	Chr	1		141	141	Filtered gender #105
108	MUS_XAN	Chr	1		142	142	Filtered gender #106
109	MUS_XAR	Chr	1		143	143	Filtered gender #107
110	MUS_XANR	Chr	1		144	144	Filtered gender #108
111	NATAM_XA	Chr	1		145	145	Filtered gender #109
112	NATAM_XAN	Chr	1		146	146	Filtered gender #110
113	NATAM_XAR	Chr	1		147	147	Filtered gender #111
114	NATAM_XANR	Chr	1		148	148	Filtered gender #112
115	PISLR_XA	Chr	1		149	149	Filtered gender #113
116	PISLR_XAN	Chr	1		150	150	Filtered gender #114
117	PISLR_XAR	Chr	1		151	151	Filtered gender #115
118	PISLR_XANR	Chr	1		152	152	Filtered gender #116
119	ROA_XA	Chr	1		153	153	Filtered gender #117
120	ROA_XAN	Chr	1		154	154	Filtered gender #118
121	ROA_XAR	Chr	1		155	155	Filtered gender #119
122	ROA_XANR	Chr	1		156	156	Filtered gender #120
123	SCAND_XA	Chr	1		157	157	Filtered gender #121
124	SCAND_XAN	Chr	1		158	158	Filtered gender #122
125	SCAND_XAR	Chr	1		159	159	Filtered gender #123
126	SCAND_XANR	Chr	1		160	160	Filtered gender #124
127	SLA_XA	Chr	1		161	161	Filtered gender #125
128	SLA_XAN	Chr	1		162	162	Filtered gender #126
129	SLA_XAR	Chr	1		163	163	Filtered gender #127
130	SLA_XANR	Chr	1		164	164	Filtered gender #128
131	CYM_XA	Chr	1		165	165	Filtered gender #129
132	CYM_XAN	Chr	1		166	166	Filtered gender #130
133	CYM_XAR	Chr	1		167	167	Filtered gender #131
134	CYM_XANR	Chr	1		168	168	Filtered gender #132
135	WEST_XA	Chr	1		169	169	Filtered gender #133
136	WEST_XAN	Chr	1		170	170	Filtered gender #134
137	WEST_XAR	Chr	1		171	171	Filtered gender #135
138	WEST_XANR	Chr	1		172	172	Filtered gender #136
139	NWEST_XA	Chr	1		173	173	Filtered gender #137
140	NWEST_XAN	Chr	1		174	174	Filtered gender #138
141	NWEST_XAR	Chr	1		175	175	Filtered gender #139
142	NWEST_XANR	Chr	1		176	176	Filtered gender #140

USING THE MAIN GENDER DATABASE

The main gender database lists given names and nicknames along with their associated gender. It is located at:

- pdGender_Main.CSV [or .TXT, .DBF]

For a large majority, the language or languages of origin and use have also been researched and included, and names from the United States provide a national popularity ranking.

PEACOCK_ID FIELD

FIELDS

PEACOCK_ID | Unique identification number (primary key)

Each record has a 17-character alphanumeric primary key that uniquely distinguishes it from all other records in the table.

The first field in the names database is PEACOCK_ID. It provides a unique primary key identifier for each record. Each begins with the character “g” to identify the software product.

The following is the first PEACOCK_ID in the main gender database:

- **g0000001-001-001F** is a complete PEACOCK_ID; no other record has this same exact identification

This field can be directly linked with *pdNickname 3.0* names database. The same identification numbers are used in both products, the only difference is the first character.

NAME FIELD

FIELDS

NAME | Stylized first name spelling

Each record has an up to 35-character alphabetic names that provide the stylized spelling of the name. All names are in UPPER CASE.

Each record lists the names that are searched for gender in the NAME field. All are in UPPER CASE.

The names in *pdGender 3.0* are the same names that are in *pdNickname 3.0* and the databases can be linked using the primary key in both files. The only difference in the primary keys is the first character which identifies the source file.

The same name can entered in more than one records stylized in different ways. For example, “Mary-Rose”, “Mary Rose”, and “Maryrose” are all the same name with hyphens and blank spaces used differently. This allows searches to more easily match the names in the gender database.

STANDARDIZING NAMES (RECOMMENDED)

While providing names in multiple stylings improves match rates, it is recommended, but not required, that users standardize their search strings before processing names to make the name data even easier to match. This involves removing all spaces, periods, hyphens, and apostrophes. Most database systems have simple commands to accomplish this. Most names, particularly English and Americanized names, do not require standardization. Standardization transforms the following characters:

From	To	Description
-	*	Hyphen
.	*	Period

* = Removed

From	To	Description
'	*	Left single quote/apostrophe
'	*	Right single quote/apostrophe

* = Removed

From	To	Description
'	*	Apostrophe
	*	Blank space

* = Removed

NORMALIZING NAMES (OPTIONAL)

Some names have grave accents, acute accents, umlauts, and other values in the extended ANSI character set. The database handles these names by providing them both with the special characters intact and normalized with all non-English alphabetic letters and glyphs converted to English A—Z letters. Additionally, because names starting with a “SAINT” prefix can also be written with as “STE” or “ST”, these names are provided in all three ways.

Users do not need to normalize name information to match against the gender files, but if users decide to make the changes themselves, normalization transforms the following characters:

From	To	Description
À	A	A-grave
Á	A	A-acute
Â	A	A-circumflex
Ã	A	A-tilde
Ä	A	A-diaeresis (umlaut)
Å	A	A-ring
Æ	AE	Æsc (grapheme)
Ç	C	C-cedilla
Ð	D	Eth
È	E	E-grave
É	E	E-acute

From	To	Description
Ì	I	I-grave
Í	I	I-acute
Î	I	I-circumflex
Ï	I	I-diaeresis (umlaut)
Ñ	N	N-tilde
Ò	O	O-grave
Ó	O	O-acute
Ô	O	O-circumflex
Õ	O	O-tilde
Ö	O	O-diaeresis (umlaut)
Ø	O	Ø-vowel

From	To	Description
ß	SS	Eszett/Sharp S
Ù	U	U-grave
Ú	U	U-acute
Û	U	U-circumflex
Ü	U	U-diaeresis (umlaut)
Ý	Y	Y-acute
Þ	Y TH P	Porn (Thorn) ^b
Ž	Z	Z-caron (grapheme)
¸	*	Spaced cedilla
˘	*	Spaced grave accent
˙	*	Spaced acute accent

Ê	E	E-circumflex	Œ	OE	Œ (grapheme)	SAINT	ST	SAINT-prefix***
Ë	E	E-diaeresis (umlaut)	Š	S SH	S-caron (grapheme)**	STE	ST	STE-prefix***

* = Removed

** The Š (S-caron) grapheme can be interpreted in two ways, as an “S” or “SH”.

*** The SAINT/STE to ST transformations should only be performed when they are being used as prefixes and are followed by a period, hyphen, or space.

^p The Þorn (Thorn) can be interpreted in three ways. Some transform it to a “P” because it obviously looks a lot like the letter. But in the days of the Anglo-Saxons it was used to pronounce what the Norman French would later introduce as the digraph “TH”. However, after the invention of the printing press, parts of the Þorn were abbreviated or dropped to the point it resembled a “Y”, which is the reason for such articulations as “Ye Olde”. In fact, the “Ye” pronunciation is still used informally today in Hiberno-English (Irish English).

EXAMPLES

The following are examples of stylized, standardized, and normalized names. Note when the stylized, standardized, and normalized names are all different, all the same, and when the stylized name is different but the standardized and normalized names the same:

Normalized	Standardized	Stylized	Gender	Language
ABDULAZIZ	ABDULAZIZ	ABDUL 'AZIZ	Male	Arabic
ANAMARIA	ANAMARÍA	ANA MARÍA	Female	Spanish
FERNANDJOSE	FERNANDJOSÉ	FERNANDO JOSÉ	Male	Portuguese, Spanish
FREDERICK	FRÉDÉRIK	FRÉDÉRIK	Male	French
JOYCE	JOYCE	JOYCE	Female	English
KATHERINE	KATHÉRINE	KATHÉRINE	Female	Swedish
MARYLOUISE	MARYLOUISE	MARY LOUISE	Female	English
ROGNVALD	RÖGNVALD	RÖGNVALD	Male	Icelandic
STGRELLAN	SAINTGRELLAN	SAINT GRELLAN	Male	Old Irish
SANDEEPKUMARA	SANDEEPKUMARA	SANDEEP KUMARA	Male	Hindi
YIGYEONG	YIGYEONG	YI-GYEONG	Female	South Korean
ZYRIEL	ZYRIEL	ZYRIEL	Female	West Pacific Filipino Tagalog

GENDER

FIELDS

GENDER | Male or female gender flag

M = Male

F = Female

Each record has a one-character alphabetic codes that indicate the gender associated with the name.

Base gender identification for each name is provided in the GENDER field. All records resolve to either male (“M”) or female (“F”) and never to a unisex name. If a name is unisex, both a male and female record is included.

This field is not intended as a gender coding field, rather it is the base information all the gender coding fields are derived from.

GIVEN AND NICK FIELDS

FIELDS

GIVEN | Given name flag*Each record has a one-character alphabetic code that indicates if the name is a given name:*

G = Given

Blank = Not a given name

NICK | Nickname flag*Each record has a one-character alphabetic code that indicates if the name is a nickname:*

N = Nickname

Blank = Not a nickname

The database identifies names as given names and nicknames. Given names are flagged with a “G” in the GIVEN field, while nicknames are flagged with an “N” in the NICK field.

Names can be flagged both as a given name and as a nickname. It is not uncommon for a nickname over time to be accepted independently as a given name. For example, “Kate”, a nickname for “Katherine” and “Katarina”, is now also considered a proper given name.

The database contains all the first name spellings gathered and published by the U.S. Census Bureau and Social Security Administration between 1800 and the present time, related nicknames, and ethnic given names and nicknames not found in the United States. About 75 percent of the given names and nicknames can be found in the United States, and the remainder only found outside the United States.

About 42% of the names in the names database are classified as given names; about 56% are classified as nicknames; and about 2% are classified as both a given name and a nickname.

Given names and nicknames are defined as follows:

GIVEN NAMES

These are formal first names (also known as personal name and forenames), normally bestowed upon, or given by parents, at or near the time of birth. This contrasts to surnames which are normally inherited and shared with other immediate family members. There are frequently numerous variations of the same given name often in multiple languages. In addition to regular given names, the names database also provides translations from other languages, and transcriptions, which are variations spelled phonetically as they sound to the person hearing and transcribing the name.

NICKNAMES

These are substitutes for the formal given names that express familiarity or endearment. For example, “Mike” is a nickname for the given name “Michael”. Nicknames can include shortened forms of the given names or diminutives.

NAME RANK

FIELDS

RANK | United States name rank

Records have an up five digit numeric value indicating the rank of first names occurring in the United States 1915 to the present as published by the U.S. Social Security Administration.

The U.S. Social Security Administration, in association with the U.S. Census Bureau, publishes a list of first names occurring 5 times or more in the United States from 1880 to the present. The names database includes all of these names and numerically ranks those with entries occurring from 1915 to the present in order of popularity, starting with the male name “James” in the #1 position, with 4,855,084 entries; and ending with 14,752 names all ranked #88,434, with five entries each. Names not occurring since 1915 or not in the Social Security data set are not ranked.

ARCHAIC NAMES

FIELDS

ARCHAIC | Archaic flag

Each record has a one-character alphabetic code that indicates if the first name is archaic and no longer in use:

A = Archaic name

Blank = Not an archaic name

About 0.8 percent of the names in the names database are archaic. They are included because of their onomastic significance. Archaic names are flagged with an “A” in the ARCHAIC field. Note that names can be archaic for one gender and in use for the opposite gender.

LANGUAGE

FIELDS

LANGUAGE | Language string

* = Rare Usage

Each record has an up to 254-character alphabetic list that indicates the language or languages of origin and use of the first name. Multiple languages are entered as a comma-delimited list with the languages in alphabetical order. Languages followed by an asterisk (for example, “English”) indicate rare usage of the name for the language and gender.*

The language or languages of origin and use are identified in the LANGUAGE field. If there is more than one language, they are listed alphabetically in a comma delimited string; for example, “Basque, Catalan, Portuguese, Spanish”.

None of the languages were derived algorithmically and the provided information represents years of extensive onomastic research. If the language has not been identified but is found in U.S. Social Security Administration or U.S. Census Bureau listings, it is flagged as “American”. When different sources list different origins and usages they may be combined depending on the reliability of the source and the reasonability of the information. The languages apply to the stylized name. Differently styled names can have different language values.

Language coverage is extensive. The list exceeds 500 languages, language families, and dialects. Some languages refer to ethnic groups. For example, Bosniak Bosnian refers to a South Slavic Muslim ethnic group inhabiting mainly Bosnia and Herzegovina.

RARE USAGES

One of the most useful features of the software is rare usages of names are identified by language. If a language is followed by an asterisk (“*”) it means the name is rare in the language and the gender flagged in the GENDER field. For example, if the language string is “Czech, Russian*, Slovak”, then “Russian” is the rare usage.

These flags are applied only to unisex names. They show when a name is used less than 20 percent of the time in the cited language and gender when precise statistics are known, or is otherwise determined to be in low usage. This indicator allows filtering out rare usages in gender coding.

TOP 30 LANGUAGES

These following are the top 30 languages, out of a total of more than 500 languages, with the number of occurrences in the names database out of a total of 397,000 names. The language count is one for each unique name formation:

1. English	225,000	11. Spanish	3,300	21. Czech	2,000
2. Arabic	46,700	12. Italian	3,100	22. Russian	1,900
3. Turkish	6,700	13. Bengali	3,100	23. Dutch	1,800
4. Punjabi	6,700	14. German	3,100	24. Hungarian	1,800
5. French	5,900	15. Pashto	3,000	25. Portuguese	1,700
6. Iranian	5,800	16. Norwegian	3,000	26. Malaysian Malay	1,700
7. Urdu	4,900	17. Danish	2,900	27. Albanian	1,700
8. Afghan Arabic	4,400	18. Korean	2,700	28. Japanese	1,700
9. Swedish	4,100	19. Egyptian Arabic	2,400	29. Bosniak Bosnian	1,500
10. Finnish	3,600	20. Polish	2,000	30. Icelandic	1,400

Note that the counts are rounded to the lower 100.

Also note that the Arabic and Muslim name section is very large due to the many different variations and ways of writing these names. These include theophoric combination names such as those with the religious prefix “Abdul”. Both common and uncommon possibilities are included, and the use of Sun Letters in Arabic and Maltese is accounted for.

A list of all the identified languages with counts is included with the software as a Microsoft Excel (XLSX) file. The language names chosen are detailed and easy to search for. For example, to select all 40 African languages, search for “Africa” in the language string, and exclude “African American”. To select all 18 Pacific Island languages, search for “Pacific” and “Oceania”. To select all 12 Indian languages, search for “Indian”, “Hindi”, and “Urdu”. To select all 38 Native American languages, search for “Native American”. Many other useful query words exist which can be determined from the provided list.

A SHORT HISTORY OF FIRST NAMES

First names have been with us for millennia, much longer than last names, which are a relatively modern development. Some of the names in the names database originated during Antiquity while others arose during the subsequent Middle Ages or during modern times. The following is a primer on recognizing these origins and the periods they represent:

ANTIQUITY

Important ancient languages include:

- **Ancient Egyptian:** attested from 3400 BC making it one of the earliest known written languages (along with Sumerian)
- **Sumerian:** the language of ancient Sumer, spoken in southern Mesopotamia (modern Iraq) and closely related to Akkadian, it is attested from 3350 BC making it one of the earliest known written languages (along with Egyptian); it was slowly replaced by Akkadian between the 3rd and 2nd millennia BC, but continued as a classical language until about 100 AD
- **Akkadian:** spoken in ancient Mesopotamia from the 29th through 8th centuries BC, including during the Akkadian Empire (ca. 2334–2193 BC), it is closely related to and replaced Sumerian, and is the earliest attested Semitic language; academic and liturgical use continued until about 100 AD
- **Assyrian:** a dialect of Akkadian spoken in upper-Mesopotamia from about 25th century BC
- **Old Aramaic:** the earliest stages of Aramaic, a Northwest Semitic language subfamily which includes Hebrew and Phoenician, dating from the 10th century BC
- **Avestan:** an Iranian language dating from the Late Bronze Age (1570–1200 BC) known only from its use as the language of Zoroastrian scripture
- **Greek:** spoken on the Balkan Peninsula since the 3rd millennium BC, and the oldest recorded living language, its earliest attested written evidence is the Linear B clay tablet found in Messenia which dates to between 1450 and 1350 BC
- **Hebrew:** a West Semitic language, closely related to Phoenician, historically regarded as the tongue of the Israelites (meaning, “Children [or Sons] of Israel”; its earliest attested written evidence, in form of primitive drawings, dates from the 10th century BC; it was nearly extinct as a spoken language by late Antiquity, but continued to be used as a literary language and as the liturgical language of Judaism, until its revival as a spoken language in the late 19th century
- **Phoenician:** a Northwest Semitic language, closely related to Hebrew, originally spoken in the ancient coastal Mediterranean region of Canaan (roughly corresponding to the Levant) and attested from the 10th until the early 4th century BC
- **Etruscan:** spoken by Etruscan civilization (768—264 BC), in Italy, in the ancient region of Etruria (modern Tuscany plus western Umbria and northern Latium) and in parts of Campania, Lombardy, Veneto, and Emilia-Romagna (where the Etruscans were displaced by the Gauls); it influenced Latin, but was eventually superseded by it.
- **Roman:** spoke Archaic Latin during the Roman Kingdom (753—509 BC) through most the Roman Republic (509—27 BC), replaced by Classical Latin around 75 BC; due to Roman conquests, Latin spread to many

Mediterranean and some northern European regions; although considered a “dead” language, Latin is still used in the creation of new words and names in modern languages

- **Koine Greek:** a dialect of Greek spoken in the Eastern Roman Empire from 300 BC to 300 AD; it is also known as Alexandrian dialect, common Attic, and Hellenistic Greek
- **Illyrian:** a family of languages spoken in the western part of the Balkans by a group of Indo-European tribes called the Illyrians; it is attested from about 500 BC
- **Alanic:** spoken by Iranian nomadic pastoral people known as the Alans from the 1st century AD
- **Gallo-Roman:** spoken by the Gauls under provincial rule in the Roman Empire from the 1st century BC to the 5th century AD
- **Proto-Germanic:** dating to the Nordic Bronze Age in Scandinavia (ca. 1700–500 BC) through Antiquity
- **Proto-Celtic:** dating from the British Iron Age (ca. 600 BC—100 AD) through Antiquity
- **Proto-Norse:** a northern dialect of Proto-Germanic from the 2nd century AD until it evolved into Old Norse at the beginning of the Viking Age about 800 AD
- **Primitive Irish:** the oldest known Goidelic language from around the 4th century AD until it evolved into Old Irish about 600 AD; it is only known from fragments, mostly personal names, inscribed on stone in the Ogham alphabet in Ireland and western Great Britain

Names from languages prefixed with “Proto-” are reconstructed and are generally unattested in any documented form.

Late Greek and Late Roman names date from late Antiquity and the early Byzantine period. Late Antiquity is generally considered from the end of the Roman Empire’s crisis of the 3rd century (ca. 235–284) to the re-organization of the Eastern Roman Empire under Byzantine Emperor Heraclius and the Islamic conquests during the early and mid 7th century.

Coptic Egyptian is the later stages of the Egyptian language spoken from the 2nd until the 17th century. Today Egyptians mainly speak a dialect of Modern Standard Arabic. Coptic Egyptian is still used as the liturgical language of the Coptic Church.

MIDDLE AGES

Almost all historians agree the Middle Ages began when the political structure of Western Europe changed at the end of the united Roman Empire (476 AD). In the database names dating from the Early Middle Ages (which followed the decline of the Western Roman Empire and is sometimes called the Dark Ages due to the relative scarcity of literary and cultural output during most of the era) are usually prefixed with “Old” such as Old High German and Old Spanish (which still continues as a liturgical language but with a modernized pronunciation). An exception is Old English which is labeled “Anglo-Saxon”. Another exception is Old Aramaic which is ancient.

Many languages went through significant changes during the High Middle Ages (a period of rapid population growth and social and political change in Europe from about the 11th through the 13th century) or by the Late Middle Ages (when prosperity and growth in Europe came to a halt and the population experienced a series of famines and plagues). Languages developing in this period are prefixed with “Middle”, or in some cases “Medieval” depending on the accepted terminology.

After Duke William II of Normandy conquered England and killed King Harold II at the Battle of Hastings (1066), the invading Normans and their descendants replaced the Anglo-Saxons as the ruling class of England. French influences were absorbed into the English language, and Old English slowly evolved into Middle English between the 12th and 15th century, additionally aided by influences from the Latin language of the church and the invention of the printing press. Nevertheless, Old English was still used throughout the Plantagenet era (1154–1485), a few years beyond the time Constantinople was finally captured by the Ottoman Turks marking the final end of the Roman Empire (1453), the conclusion of the Middle Ages in the minds of many historians. Of course others cite the Battle of Bosworth Field which established the Tudor dynasty and an era of expansion for England (1485), the conquest of Granada and its annexation by Castile ending Islamic rule (1492), the discovery of the Americas by Christopher Columbus (also 1492), the death of Queen Isabella I of Castile (1504), the death of her spouse King Ferdinand II of Aragon (1516), and the Protestant Reformation (1517) as more appropriate cutoff points, often influenced by the nationality of the historian.

There was no similar revision during the High or Late Middle Ages in many languages, including Spanish, and they do not have a generally recognized middle variety.

Tiberian Hebrew is the canonical pronunciation of the Hebrew Bible (or Tanakh) committed to writing by Masoretic scholars living in the Jewish community of Tiberias in ancient Palestine (ca. 750–950). It is written in a form of Tiberian vocalization dating from the 8th century, but the oral tradition it reflects has ancient roots. Tiberian pronunciation of Hebrew is considered by textual scholars to be the most exact and proper pronunciation of the language as it preserves the original Semitic consonantal and vowel sounds of ancient Hebrew.

Much is unknown about the origin of the Yiddish, a High German language written in the Hebrew alphabet, because most speakers were exterminated in the Holocaust. The consensus among scholars is it emerged among the Ashkenazi Jews in Central Europe between the 10th and 12th centuries and later spread to Eastern Europe in the 16th century.

MODERN

Language formations after the Middle Ages are usually know as modern.

There is frequently confusion about the development of the three modern strains of Gaelic: Irish, Scottish, and Manx. All three sprang from Middle Irish which came from Old Irish.

SPECIAL AND UNIQUE ORIGINS

FIELDS

SPECIAL | Special origin string

Records have an up to 100-character alphabetic list that indicates special and unique characteristics about the origin of the name. Multiple characteristics are entered as a comma-delimited list with the elements in alphabetical order. This field is also used as a flow-over field for the language string.

Many records provide information about special and unique origins. The special and unique characteristics of origin are identified in the SPECIAL field. If there is more than one element, they are listed alphabetically in a comma delimited string; for example, “Biblical, Latinized Greek Mythology, Roman cognomen, Surname”.

Additionally, this field is also used as a flow-over field for the language string when the number of languages exceeds the LANGUAGE field length. In the handful of cases this happens, the additional languages are entered at the beginning of the SPECIAL field following a plus (“+”) sign. If there are also special origins, they are entered after the languages following a semicolon (“;”); for example, “+ Slovene, Swedish; Biblical, Greek byname, Surname”.

Special and unique origins include:

- Names from religion; identifications are:
 - **Biblical**
 - **Quranic**
 - **Sanskrit**
- Bynames: a familiar name for a person, similar to a nickname, that is often used as a replacement for a personal name—for example, Rocky is a common byname for boxers; identifications are:
 - **American English byname**
 - **Ancient Germanic byname**
 - **Anglo-Saxon byname**
 - **English byname**
 - **French byname**
 - **German byname**
 - **Greek byname**
 - **Italian byname**
 - **Middle English byname**
 - **Middle High German byname**
 - **Old French byname**
 - **Old High German byname**
 - **Old Irish byname**
 - **Old Norman French byname**
 - **Old Norse byname**
 - **Old Welsh byname**
 - **Polish byname**

- **Roman byname**
 - **Welsh byname**
- **History** = Names that became known through historical events
- **Literature** = Literary names created by authors, composers, and poets
- Names from mythology and legend; identifications are:
 - **Anglicized Egyptian Mythology**
 - **Anglicized Greek Mythology**
 - **Anglicized Judeo-Christian Legend**
 - **Anglicized Roman Mythology**
 - **Anglicized Welsh Arthurian Legend**
 - **Anglo-Saxon Arthurian Legend**
 - **Anglo-Saxon Mythology**
 - **Baltic Mythology**
 - **Breton Mythology**
 - **Celtic Mythology**
 - **Egyptian Mythology**
 - **English Arthurian Legend**
 - **European Mythology**
 - **Far Eastern Mythology**
 - **Finnish Mythology**
 - **French Arthurian Legend**
 - **German Arthurian Legend**
 - **Germanic Mythology**
 - **Greek Mythology**
 - **Hebrew Mythology**
 - **Hellenized Egyptian Mythology**
 - **Hellenized Near Eastern Mythology**
 - **Hellenized Persian Mythology**
 - **Hindu Mythology**
 - **Irish Mythology**
 - **Islamic Mythology**
 - **Judæo-Christian Legend**
 - **Latinized Egyptian Mythology**
 - **Latinized Germanic Mythology**
 - **Latinized Greek Mythology**
 - **Latinized Irish Mythology**
 - **Latinized Near Eastern Mythology**
 - **Mayan Mythology**
 - **Near Eastern Mythology**
 - **Norse Mythology**
 - **Northwest Semitic Mythology**
 - **Orphic Mythology**
 - **Persian Mythology**

- **Proto-Germanic Mythology**
- **Roman Mythology**
- **Russian Mythology**
- **Scottish Arthurian Legend**
- **Slavic Mythology**
- **Spanish Arthurian Legend**
- **Welsh Arthurian Legend**
- **Welsh Mythology**
- **Roman cognomen** = (plural cognominia) were originally nicknames that were later utilized to augment family names to identify a particular branch within a family or family within a clan
- **Roman gens** = (plural gentes) identified a family consisting of all those individuals who shared the same nomen and claimed descent from a common ancestor
- **Roman nomen** = (plural nomina) were hereditary surnames that identified a person as a member of a distinct gens
- **Roman praenomen** = (plural praenomina) are early personal names chosen by the parents of a Roman child originally bestowed the eighth day after the birth of a girl, or the ninth day after the birth of a boy; the praenomen would then be formally conferred a second time when girls married, or when boys reaching manhood and assumed the toga virilis (which in the case of Romans boys was about age 14 or 15)
- **Surname** = Name is also a surname in the *pdSurname* companion software product

WORLD FIELD

FIELDS

WORLD | Unfiltered international male or female gender flag

M = Male

F = Female

U = Unisex

Each record has a one-character alphabetic codes that indicate the unfiltered gender associated with the name.

The WORLD field is the first in a series of 141 gender coding fields. Notably, it is the only gender coding field without filters of any kind. It is called “world” because it defines the basic international usage of each name. It can be utilized like the standard unfiltered gender coding fields most users are familiar with in other products.

This field identifies names as male (“M”), female (“F”) or, when the name is both male and female, unisex (“U”). It derives the largest number of unisex identifications because it gives equal weight to all languages and nationalities. If a name is male in the United States and female in Vietnam, this field will flag the name as unisex.

FILTERED GENDER CODING FIELDS

FIELDS

WORLD_XA, WORLD_XAN, WORLD_XAR, WORLD_XANR | International filters
USA_XA, USA_XAN, USA_XAR, USA_XANR | Unites States filters
US_ES_XA, US_ES_XAN, US_ES_XAR, US_ES_XANR | United States/Spanish (Español) filters
US_HS_XA, US_HS_XAN, US_HS_XAR, US_HS_XANR | United States/Hispanic filters
US_FR_XA, US_FR_XAN, US_FR_XAR, US_FR_XANR | United States/French filters
ENG_XA, ENG_XAN, ENG_XAR, ENG_XANR | English filters
EN_AA_XA, EN_AA_XAN, EN_AA_XAR, EN_AA_XANR | **English/African American filters**
EN_ES_XA, EN_ES_XAN, EN_ES_XAR, EN_ES_XANR | English/Spanish (Español) filters
EN_HS_XA, EN_HS_XAN, EN_HS_XAR, EN_HS_XANR | English/Hispanic filters
EN_FR_XA, EN_FR_XAN, EN_FR_XAR, EN_FR_XANR | English/French filters
AFRAM_XA, AFRAM_XAN, AFRAM_XAR, AFRAM_XANR | African American filters
SPA_XA, SPA_XAN, SPA_XAR, SPA_XANR | Spanish filters
HISP_XA, HISP_XAN, HISP_XAR, HISP_XANR | Hispanic filters
FRA_XA, FRA_XAN, FRA_XAR, FRA_XANR | French filters
AFR_XA, AFR_XAN, AFR_XAR, AFR_XANR | African (non-Muslim) filters
BRIT_XA, BRIT_XAN, BRIT_XAR, BRIT_XANR | British filters
CEL_XA, CEL_XAN, CEL_XAR, CEL_XANR | Celtic (language family) filters
EASIA_XA, EASIA_XAN, EASIA_XAR, EASIA_XANR | East Asian filters
EA_PI_XA, EA_PI_XAN, EA_PI_XAR, EA_PI_XANR | East Asian/Pacific Islander filters
GAEL_XA, GAEL_XAN, GAEL_XAR, GAEL_XANR | Gaelic (Goidelic language family) filters
DEU_XA, DEU_XAN, DEU_XAR, DEU_XANR | German filters
GEM_XA, GEM_XAN, GEM_XAR, GEM_XANR | Germanic (language family) filters
HAW_XA, HAW_XAN, HAW_XAR, HAW_XANR | Oceania Hawaiian filters
IND_XA, IND_XAN, IND_XAR, IND_XANR | Indian (South Asia) filters
ITA_XA, ITA_XAN, ITA_XAR, ITA_XANR | Italian filters
JW_XA, JW_XAN, JW_XAR, JW_XANR | Jewish/Yiddish/ Hebrew filters
MUS_XA, MUS_XAN, MUS_XAR, MUS_XANR | Muslim filters
NATAM_XA, NATAM_XAN, NATAM_XAR, NATAM_XANR | Native American filters
PISLR_XA, PISLR_XAN, PISLR_XAR, PISLR_XANR | Pacific Islander filters
ROA_XA, ROA_XAN, ROA_XAR, ROA_XANR | Romance (language family) filters
SCAND_XA, SCAND_XAN, SCAND_XAR, SCAND_XANR | Scandinavian filters
SLA_XA, SLA_XAN, SLA_XAR, SLA_XANR | Slavic filters
CYM_XA, CYM_XAN, CYM_XAR, CYM_XANR | Welsh (Cymraeg) filters
WEST_XA, WEST_XAN, WEST_XAR, WEST_XANR | Western World (Occident) filters
NONW_XA, NONW_XAN, NONW_XAR, NONW_XANR | Non-Western World filters

M = Male
 F = Female
 U = Unisex

Each record has 140 one-character alphabetic codes that indicate filtered genders associated with the name.

The gender database has a series of 140 filtered gender coding fields which are the heart of the *pdGender* matching system. They allow filtering the gender coding output for languages and nationalities, rare usage by one gender, archaic names, and nicknames. Here are examples:

- When names are different genders in different languages and nationalities, users can choose which languages and nationalities to take precedence
- When names are one gender now in current times, but were the opposite gender in a previous era, the system automatically applies the modern usage
- When names are one gender when used as a proper given name and unisex when used as a nickname, users can choose to have the given name usage applied
- When unisex names are only rarely used by one gender and much more common in the opposite gender, users can choose to ignore the rare instances

The field names are designed to indicate what filters are applied. Here are examples:

	Field name	Description
<i>Example 1</i>	WORLD_XA	Gives equal weight to all languages and nationalities and reduces precedence of archaic names in gender determination
<i>Example 2</i>	USA_XAN	Prioritizes United States names and reduces precedence of archaic names and nicknames in gender determination
<i>Example 3</i>	EN_FR_XAR	Prioritizes English and French names and reduces precedence of archaic names and rare usages in gender determination
<i>Example 4</i>	HISP_XANR	Prioritizes Hispanic names and reduces precedence of archaic names, nicknames, and rare usages in gender determination

PRIORITIZING LANGUAGES AND NATIONALITIES

Because names can have different genders in different languages and nationalities, a filter is provided allowing the choice of which languages and nationalities to take precedence. There are 35 options which are indicated in the prefix of each gender coding field name. The choices and field name prefixes are:

Prefix	Filter
WORLD_	All languages and nationalities are given equal weight
USA_	United States names are prioritized
US_ES_	United States and Spanish (Español) names are prioritized
US_HS_	United States and Hispanic names are prioritized
US_FR_	United States and French names are prioritized
ENG_	English names are prioritized
EN_AA_	English and African American names are prioritized
EN_ES_	English and Spanish (Español) names are prioritized
EN_HS_	English and Hispanic names are prioritized
EN_FR_	English and French names are prioritized
AFRAM_	African American names are prioritized
SPA_	Spanish names are prioritized
HISP_	Hispanic names are prioritized
FRA_	French names are prioritized
AFR_	African (non-Muslim) names are prioritized
BRIT_	British names are prioritized

<i>CEL_</i>	Celtic (language family) names are prioritized
<i>EASIA_</i>	East Asian names are prioritized
<i>EA_PI_</i>	East Asian and Pacific Islander names are prioritized
<i>GAEL_</i>	Gaelic (Goidelic language family) names are prioritized
<i>DEU_</i>	German (Deutsch) names are prioritized
<i>GEM_</i>	Germanic (language family) names are prioritized
<i>HAW_</i>	Oceania Hawaiian names are prioritized
<i>IND_</i>	Indian (South Asia) names are prioritized
<i>ITA_</i>	Italian names are prioritized
<i>JW_</i>	Jewish, Yiddish, and Hebrew names are prioritized
<i>MUS_</i>	Muslim names are prioritized
<i>NATAM_</i>	Native American names are prioritized
<i>PISLR_</i>	Pacific Islander names are prioritized
<i>ROA_</i>	Romance (language family) names are prioritized
<i>SCAND_</i>	Scandinavian names are prioritized
<i>SLA_</i>	Slavic (language family) names are prioritized
<i>CYM_</i>	Welsh (Cymraeg) names are prioritized
<i>WEST_</i>	Western World (Occident) names are prioritized
<i>NONW_</i>	Non-Western World names are prioritized

ADDING OTHER FILTERS

The suffix of each gender coding field name indicates any additional filters that are applied. They all begin with an “X”, indicating eXclusion, followed by up to three characters (“A”, “N”, and/or “R”, in respective order) showing what filters are applied. There are four possible suffixes:

Suffix	Filter
<i>_XA</i>	Reduces precedence of archaic names in gender determination
<i>_XAN</i>	Reduces precedence of archaic names and nicknames in gender determination
<i>_XAR</i>	Reduces precedence of archaic names and rare usages in gender determination
<i>_XANR</i>	Reduces precedence of archaic names, nicknames, and rare usages in gender determination

A (ARCHAIC FILTER)

When an “A” is in the suffix (such as “_XA”, “_XAN”, “_XAR”, and “_XANR”) it means archaic usage is ignored in gender determination. This is useful because some names change gender over the years, and this filter insures the modern usage is applied. This filter is applied automatically.

N (NICKNAME FILTER)

When an “N” is in the suffix (such as “_XAN” and “_XANR”) it means nickname usage is reduced in priority in gender determination. This is useful because often names are one gender when used as a proper given name and unisex when used as a nickname, and this optional filter can insure the given name usage is applied in these instances.

R (RARE USAGES BY ONE GENDER FILTER)

When an “R” is in the suffix (such as “_XAR” and “_XANR”) it means rare usage is reduced in priority in gender determination. This is useful because often unisex names are only rarely used by one gender and much more common in the opposite gender, and this optional filter can insure that the common usage (when more than 80% usage is by one gender) is applied in these instances.

FUZZY LOGIC (PRO ONLY)

This section applies to pdGender Pro only.

If you typed “**Garfeild**” into a word processor, it would probably be underlined with a squiggly red line signifying a misspelling. It is the name “**Garfield**” with the “**IE**” reversed to “**EI**”—a common mistake.

The fuzzy logic technology in the *Pro* edition of this software allows gender coding name data that has typographical errors. If users look at the fuzzy logic records, they are likely to see errors they have repeatedly made or seen. In many cases you will have to look close to see the difference, but they are different. There are more than 3.5 million fuzzy logic records.

USING THE FUZZY LOGIC FILE

The fuzzy logic file is utilized exactly like the main gender coding file, except fuzzy logic deals with misspelled names instead of real given names and nicknames. It has the same NAME field to search in (which in this case contain the misspelled name spellings), the same WORLD unfiltered gender coding field, and the same 140 filtered gender coding fields. It does not have the remaining fields because they are not pertinent for this file. The gender provided in each gender coding field is the most likely for each misspelling and selected filter. The fuzzy logic file is located at:

- pdGender_FuzzyLogic.CSV [or .TXT, .DBF]

Fuzzy logic attempts to duplicate real errors created while entering names into databases. The most likely typographical errors are determined based on the number of letters, the characters involved, where they are located in the name, the language, and other factors. None of the fuzzy spellings formulate a real name already in the database.

The biggest advantage in our technology is in its ability to work with language rules that indicate how individuals of various nationalities may hear and spell names.

Some fuzzy logic spellings have one typographical error while others have multiple issues, so the technology is suited for even the worst typists and transcribers.

The algorithms have five layers:

PHONETIC MISSPELLINGS

These algorithms look at digraphs, trigraphs, tetragraphs, pentagraphs, hexagraphs, and even a German heptagraph, “SCHTSCH”, used to translate Russian words with the “SHCHA” or “SHCH” (romanticized) sound. These are, respectively, two to seven letter sequences that form one phoneme or distinct sound. Most of letter sequences trigraph and above are Irish who have more language rules than you can shake a stick at.

Many misspellings occur as transcribers enter the sounds they hear. The character sequences and the sounds they produce are different for each language and situation, such as before, after, or between certain vowels and consonants, so our substitutions are language-rule based. Furthermore, our algorithms consider both how a name may sound to someone who speaks English as well as how it may sound to someone who speaks Spanish, which is often different. Take the digraph “SC”. Before the vowels “E” or “I” it is most likely to be misspelled by an English speaker as “SHE” or “SHI” while a Spanish speaker may hear “CHE” or “CHI” and sometimes “YE” or “YI”. Our library includes over 80,000 language-based letter sequence phonetic rules. Phonetic misspelling examples:

	Real name	Fuzzy name	Gender
<i>Example 1</i>	BARTHOLOMEW	BARTHOLOMUE	Male
<i>Example 2</i>	DAWNETTE	DAUNETTE	Female
<i>Example 3</i>	NATHANIEL	NATHANAIL	Male
<i>Example 4</i>	PHYLLIS	FYLLIS	Female
<i>Example 5</i>	SIGOURNEY	SIGOURNI	Female
<i>Example 6</i>	XAVIER	XAVAR	Male

REVERSED DIGRAPHS

These algorithms look for misspellings due to reversed digraphs (two letter sequences that form one phoneme or distinct sound) which are a common typographical issue, such as “IE” substituted with “EI”. The character sequences and the sounds they produce are different for each language and situation, such as before, after, or between certain vowels and consonants, so our substitutions are language-rule based. Reversed digraph examples:

	Real name	Fuzzy name	Gender
<i>Example 7</i>	ANNABETH	ANNABEHT	Female
<i>Example 8</i>	CAETLIN	CEATLIN	Female
<i>Example 9</i>	EUGENE	UEGENE	Male
<i>Example 10</i>	FRIEDRICH	FREIDRICH	Male
<i>Example 11</i>	RAQUEL	RAUQEL	Female
<i>Example 12</i>	VICKTOR	VIKTOR	Male

DOUBLE-LETTER MISSPELLINGS

These algorithms look for misspellings due to double letters typed as single letters and single letters that are doubled. The most common typographical issues occur with the characters, in order of frequency, “SS”, “EE”, “TT”, “FF”, “LL”, “MM”, and “OO”. Double-letter misspelling examples:

	Real name	Fuzzy name	Gender
<i>Example 13</i>	EMANNUEL	EMMANNUEL	Male
<i>Example 14</i>	KASSANDREA	KASANDREA	Female

MISSED LETTERS

These algorithms look for missed keystrokes and provide fuzzy logic matches with missing letters. Unlike the other algorithms, these are not language specific. Keystrokes can be missed in any language. Missed letter examples:

	Real name	Fuzzy name	Gender
<i>Example 15</i>	ABDUL	ADUL	Male
<i>Example 16</i>	MARGARET	MRGARET	Female

STRING MANIPULATIONS

These algorithm changes letters and syllables in a variety of ways. They are less guided by language rules and more guided by randomness. String manipulation examples:

	Real name	Fuzzy name	Gender
<i>Example 17</i>	CYNTHIA	CYNTTHA	Female
<i>Example 18</i>	GERALD	GERLLD	Male

COMPATIBILITY

To ensure compatibility with any operating system and database platform, **pdGender** is provided in multiple file formats and utilizes only the ANSI character set (ASCII values 0 to 127 and extended values 128 to 255).

USING PDGENDER WITH PDNICKNAME AND PDSURNAME

pdGender, *pdNickname*, and *pdSurname* make excellent partners. They have been developed to be fully compatible. The name pair format in *pdNickname* is very similar to the *pdSurname* database except *pdNickname* is used to match give names and nicknames while *pdSurname* matches last names. *pdGender* is based on the first name database and is designed to apply gender identification to first name records. Note that *pdNickname* and *pdSurname* are not required to use *pdGender* but they are highly attuned to work together.

USER GUIDE UPDATES

User guides are updated based on information gained from user experience. It is suggested that users regularly check the Support section of the Peacock Data website for updates. Look for a date newer than the date below:

The publication date of this guide is: May 10, 2016.

DATABASE VERSION NUMBER

Depending on the file format, the version number of each copy of *pdGender* is written in the first or second row of the first or second column of all database files in **X.X.X** format. The first number is the main version number of the release. The number after the first dot is the update for the version indicated. The number after the second dot references a minor revision.

SITE LICENSE

Peacock Data's site licenses are designed to be fair. They are broader than most software licenses in that they allow installation on not one but all computers in the same building within a single company or organization. We ask users to honor these simple rules so Peacock Data can continue bringing great products to users.

THE USE OF PDGENDER IS GOVERNED BY THE FOLLOWING SITE LICENSE

- I. This Site License grants to the Licensee the right to install the licensed version of **pdGender** (hereinafter, 'information') on all computers in the same building within a single company or organization. Separate Site Licenses must be purchased for each building the information is used in.
- II. The information may only be used by the employees of the Licensee. If the Licensee is an educational institution, the data may only be used by enrolled students, faculty, teaching assistants, and administrators.
- III. Temporary employees, contractors, and consultants of the Licensee who work on-site at the Licensee's facility may also use the information in connection with the operation of the business of the Licensee. Any copies of the information used by temporary employees, contractors, and consultants must be removed from such individual's computers once they cease working at the Licensee's facility.
- IV. The information cannot be used to provide services or products to customers or other third parties, whether for-profit or given away. A Developer License must be purchased separately by the Licensee to incorporate the information in for-profit services and products.
- V. The Licensee is required to use commercially reasonable efforts to protect the information and restrict network or any other access to the information by anyone inside or outside of the Licensee's facility who is not authorized to use the information.
- VI. The Licensee owns the media on which the information is recorded or fixed, but the Licensee acknowledges that Peacock Data, Inc. and its licensors retain ownership of the information itself.
- VII. The Licensee may not transfer or assign its rights under this license to another party without Peacock Data, Inc.'s prior written consent.
- VIII. Peacock Data, Inc. may revoke the rights granted by this license upon a violation of any provision herein by the Licensee.
- IX. This Site License is governed by Peacock Data, Inc.'s Terms of Service and Privacy Policy, and the laws and regulations of the United States and the State of California.

COPYRIGHT NOTICE

pdGender is Copyright © 2009-2016 Peacock Data, Inc. All Right Reserved.