



PRACTICAL MAGIC WITH THE SOCIAL SECURITY DEATH MASTER FILE

June 21, 2012

Philip L. Adams, ASA, MAAA
North American Biometric Research



Roadmap



-
1. The Data
 2. Making the most of it with text mining
 3. Other Issues
 - Privacy
 - Escheatment
 - Regulatory Discussions


DEATH MASTER FILE

WHAT'S IN IT




The Data

-
- Fixed width text file
 - Social Security Number
 - Last name, first name, middle name, name suffix
 - Date of birth
 - Date of death
 - Verified or proof of death
 - Older records may not have this flag
 - Before November 2011, state of residence, last zip code and zip code of lump sum payment

The Data Munich RE 

- SSA does not guarantee accuracy of file
 - Being in the DMF does not mean that you are dead
 - Being absent from the DMF does not mean that you are alive
- Can be dirty
 - Permuted digits in dates, numbers
 - Permuted or improperly entered letters
 - Maiden names

Title of presentation and name of speaker 15/06/2012 5

The Data Munich RE 

Why Deaths May Not Be Listed

Death has not yet been reported	Person did not participate in Social Security	Survivor benefits are being paid to a spouse or dependent
Death is too recent	SSA has less financial incentive to collect information for non-beneficiaries	Death was reported only through state vital records

Title of presentation and name of speaker 15/06/2012 6

The Data
Reporting Rates by Age



Historical Hit Rates in DMF							
Year	Total	Age					
		0-24	25-54	55-64	65-74	75-84	85+
1980	91.9%	34.4%	92.0%	93.4%	96.5%	96.2%	97.1%
1985	89.1%	17.0%	76.4%	86.1%	96.2%	96.3%	96.5%
1990	75.3%	15.5%	50.0%	57.7%	85.0%	87.1%	85.7%
1995	87.5%	42.6%	71.4%	78.3%	93.5%	94.2%	93.4%

Hill, M. and Rosenwaike, I. The Social Security Administration's Death Master File: The Completeness of Death Reporting at Older Ages. Social Security Bulletin. 2002 v64 iss1

More recent match rates are above 99%.
(May 2011, Federal Computer Week)



DEATH MASTER FILE
HOW TO WORK WITH IT



Text Mining

-
- Exact matches are too stringent
 - Matching algorithms for names
 - Phonetic
 - String distance functions
 - Some tolerance in matching dates and SSNs
 - 06/10/2005 versus 10/06/2005 versus 06/11/2005
 - Maiden names (can underlink by as much as 5%)
 - Initials and abbreviations


Fuzzy Matching with Soundex

-
1. Retain the first letter of the name
 2. Drop all vowels and near vowels (A, E, I, O, U, Y, H, W)
 3. Replace consonants with the following rules:

Letter	Mapping
B, F, P, V	1
C, G, J, K, Q, S, X, Z	2
D, T	3
L	4
M, N	5
R	6

4. Adjacent letters with same number are encoded with single number, as are two letters separated by H or W
5. Continue until you have one letter and three numbers; pad with 0s if necessary


Fuzzy Matching with Soundex
Pathologies

Munich RE 

Case	Name	Soundex
Slight Misspelling	Philip	P410
	Phillip	P410
Similar surnames	Adams	A352
	Addams	A352
Slavic Variations	Moskowitz	M232
	Moskovitz	M213
Typos	Adams	A352
	Adasm	A325
Distinct Preimage	Adams	A352
	Addington	A352

Title of presentation and name of speaker 15/06/2012 11

Fuzzy Matching with Soundex
Pathologies

Munich RE 

- Names sharing same Soundex can be broad
- ROBERTS has the same Soundex as
 - RAFFERTY
 - RIVARD
 - ROBARDS
 - RAIFORD

Title of presentation and name of speaker 15/06/2012 12

Fuzzy Matching with Jaro-Winkler String Distance

Munich RE 

- Normalized such that 0 means no similarity and 1 means perfect match
- Jaro distance

$$d_j = \frac{1}{3} \left(\frac{m}{|s_1|} + \frac{m}{|s_2|} + \frac{m-t}{m} \right)$$

- m is the number of matching characters within

$$\left\lfloor \frac{\max(|s_1|, |s_2|)}{2} \right\rfloor - 1$$

- t is half the number of transpositions

Wikipedia article

Title of presentation and name of speaker 15/06/2012 13

Fuzzy Matching with Jaro-Winkler String Distance

Munich RE 

- Jaro-Winkler String Distance

$$d_w = d_j + lp(1 - d_j)$$

- l is the length of the common prefix up to 4 characters
- p is an arbitrary weight at most 0.25.

Title of presentation and name of speaker 15/06/2012 14

Fuzzy Matching with Jaro-Winkler String Distance

Case	Name 1	Name 2	Jaro-Winkler Distance
Slight Misspelling	Philip	Phillip	0.97
Similar surnames	Adams	Addams	0.96
Slavic Variations	Moskowitz	Moskovitz	0.96
Typos	Adams	Adasm	0.95
Distinct Preimage	Adams	Addington	0.63

Tools to Make Text Mining Easier

-
- Search for “record linkage” in your favorite internet search engine
 - Fuzzy matching
 - Oracle (via the UTL_MATCH package)
 - SQL Server (via SQL Server Integration Services)
 - R (package `RecordLinkage`)
 - ...perhaps even in your favorite application

DEATH MASTER FILE

OTHER ISSUES



Issues Surrounding Usage Privacy

-
- Deaths are a matter of public record
 - Right to privacy ends with death
 - ...except if there is an enforceable contract on the living (e.g., wills and trusts, company privacy policies, etc.)
 - ...or if a law provides for it (e.g., HIPAA and medical records)

Issues Surrounding Usage Privacy

Munich RE 

-
- Concerns about identity theft from DMF
 - Congressional actions
 - SB 1534 / HR 3215: Empowers Secretary of Commerce to develop a certification program for use of DMF; applicants must demonstrate interest in fraud prevention
 - HR 3475: Amends Social Security Act to limit use of DMF to federal and state agencies for research and tax purposes
 - Prior to November 2011, SSA improperly disclosed deaths
 - Removal of deaths where death is reported only through a state
 - Of the nearly 2.8 million deaths per year, will lose approximately 1 million
 - Geographical identifiers were also removed from all records
 - SSA suggests going to state vital records offices

Title of presentation and name of speaker 15/06/2012 19

Issues Surrounding Usage Escheatment and Regulatory Concerns

Munich RE 

-
- MetLife, NAIC and \$400 million
 - Conflicting forces
 - Pending legislation to protect against identity theft may hinder usage of DMF
 - State regulators demand life insurers check for unclaimed death benefits
 - Discussions ongoing at ACLI, NAIC
 - Strong feeling that life insurers have “legitimate” need for full access to DMF
 - Concerns that privacy legislation will hinder payments to beneficiaries (by prohibiting contact with third parties, etc.)

Title of presentation and name of speaker 15/06/2012 20

THANK YOU!

