

# Bridging the Generation Gap--Bringing Ancestors and Descendants Together with Technology

David Ouimette

25 March 2004

# Which Ancestors Can I Find Online?

Pierre Ouimette

Foussaint Ouimet

David Ouimet

Celeste Guerin

Luce Laurin

Joseph Laurin

Charles Ouimette

Agathe Desnoyers

Pierre Robert Lafontaine Joseph Robert

Cordelia Robert Lafontaine

Euphrosine Frechet

Richard Ouimette

Marie Sausse Comtois

Euphrosine Huot

Alexis Huot

Jozef Czelusniak

Grzegorz Czelusniak

Stanislaw Czelusniak

Katarzyna Krupa

| Katarzyna Trzaskos

Walenty Trzaskos

Eleanora Czelusniak

Agata Dybas

Jakub Jarek Jan Jarek

Katarzyna Jarek

Katarzyna Bujak

Anna Czelusniak

Jan Czelusniak

David Scott Ouimette Marianna Dybas

John Connor

Michael O'Connor

|

William O'Connor

Ambrose Lysaght Browne Bartholomew Browne

Mary Jane Browne

Margaret Lysaght

Mary West Thomas West

Marilyn O'Connor

Marble Bissonette Amable Bissonnet

Alfred Marble Bissonette Phebe Battiste

| Clarissa White

Newell White

Merle Carlyss Bissonett

Clarissa Mills

Edward Tallman

Joseph Tallman

Blanche Elvina Tallman

Jane Patterson

Clarissa Wade Horatio Wade

Phebe Tarte

# Records for These Ancestors Are Online

Pierre Ouimet

Toussaint Ouimet

David Ouimet

Celeste Guerin

Luc Laurin

Joseph Laurin

Charles Ouimette

Agathe Desnoyers

Pierre Robert Lafontaine Joseph Robert

Cordelia Robert Lafontaine

Euphrosine Frechet

Richard Ouimette

Marie Sausse Comtois

Euphrosine Huot

Alexis Huot

Jozef Czelusniak

Grzegorz Czelusniak

Stanislaw Czelusniak

Katarzyna Krupa

Katarzyna Trzaskos

Walenty Trzaskos

Eleanora Czelusniak

Agata Dybas

Jakub Jarek Jan Jarek

Katarzyna Jarek

Katarzyna Bujak

Anna Czelusniak

Jan Czelusniak

David Scott Ouimette Marianna Dybas

John Connor

Michael O'Connor

William O'Connor

Ambrose Lysaght Browne Bartholomew Browne

Mary Jane Browne

Margaret Lysaght

Mary West Thomas West

Marilyn O'Connor

Marble Bissonette Amable Bissonnet

Alfred Marble Bissonette Phebe Battiste

Clarissa White

Newell White

Merle Carlyss Bissonett

Clarissa Mills

Edward Tallman

Joseph Tallman

Blanche Elvina Tallman

Jane Patterson

Clarissa Wade Horatio Wade

Phebe Tarte

# These Ancestors Are Found Online or at the FHL

Pierre Ouimette

Toussaint Gummer

David Ouimet

Celeste Guerin

Luce Laurin

Joseph Laurin

Charles Ouimette

Agathe Desnoyers

Pierre Robert Lafontaine Joseph Robert

Cordelia Robert Lafontaine

Euphrosine Frechet

Richard Ouimette

Marie Sausse Comtois

Euphrosine Huot

Alexis Huot

Jozef Czelusniak

Grzegorz Czelusniak

Stanislaw Czelusniak

Katarzyna Krupa

| Katarzyna Trzaskos

Walenty Trzaskos

Eleanora Czelusniak

Agata Dybas

Jakub Jarek Jan Jarek

Katarzyna Jarek

Katarzyna Bujak

Anna Czelusniak

Jan Czelusniak

David Scott Ouimette Marianna Dybas

John Connor

Michael O'Connor

William O'Connor

Ambrose Lysaght Browne Bartholomew Browne

Mary Jane Browne

Margaret Lysaght

Mary West Thomas West

Marilyn O'Connor

Marble Bissonette Amable Bissonnet

Alfred Marble Bissonette Phebe Battiste

| Clarissa White

Newell White

Merle Carlyss Bissonett

Clarissa Mills

Edward Tallman

Joseph Tallman

Blanche Elvina Tallman

Jane Patterson

Clarissa Wade Horatio Wade

Phebe Tarte

# LDS Family and Church History Department

- Challenge:
  - To help many more members of the Church find their ancestors
- Solutions:
  - Build a Common Pedigree
  - Make family history research easier
  - Convert microfilms to digital images and indexes
  - Provide access to records from around the world

# Bridging the Gap between a Researcher and Her Ancestor

- The ancestor's records are in an archive far away
  - Microfilm the records
  - Digitize the microfilm and post the images
  - Offer access to images by locality, date range and event
  - Index the records by name
  - Normalize the name index
  - Reconstruct families from the index
  - Link indexed families to researcher's tree
- Each step of this process has several technology challenges

# Two Categories of Technology Problems

- Issues with family trees
- Issues with researching historical records

# Family History Research Problems with Family Trees

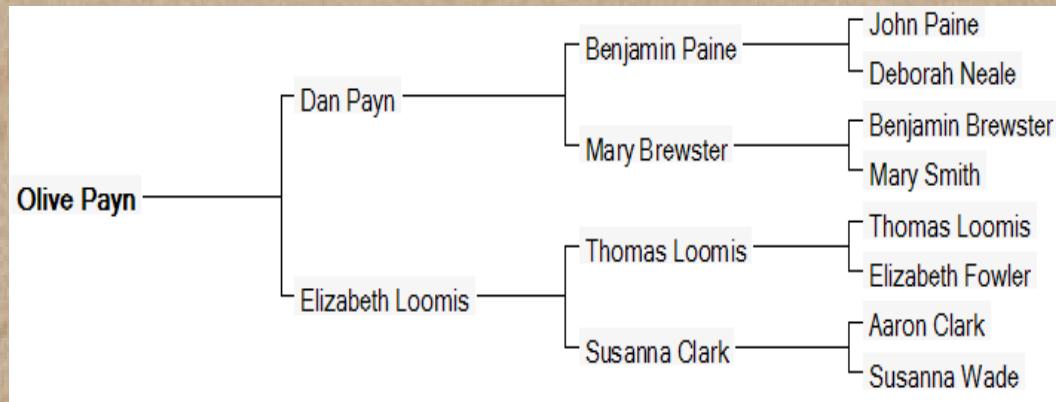
- Record linkage
- Data normalization (names, places, events)
- Research guidance
- Collaboration
- Cultural differences

# Family History Research Problems with Records and Research

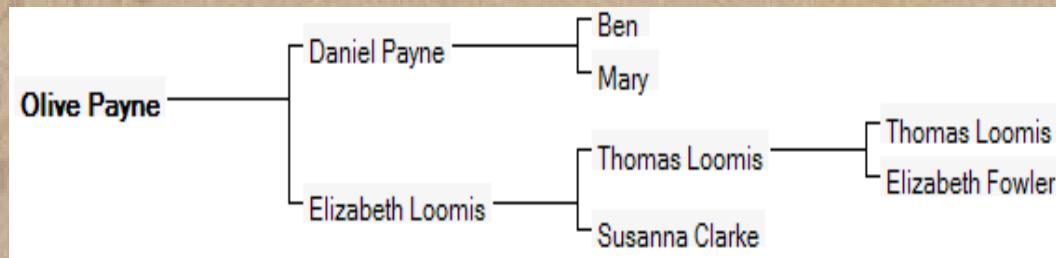
- Research model
- Digital image conversion
- Digital image enhancement
- Better transcriptions
- Handwriting recognition
- Better OCR technologies
- Metadata taxonomy
- Metadata mining
- Family reconstitution

DATE OF DEATH.			FULL NAME OF THE DECEASED.
MONTH.	DAY.	YEAR.	
June	1	1874	Capitolia Austin
May	9	1874	John B. Truitt
Nov.	10	1874	Clarance Truitt
Oct.	20	1874	Francis Paulist
Jan.	25	1874	Theolacia Whitelaw
Jan.	9	1874	Gland Gane
Dec.	18	1874	Hamiland Thayw.
Nov.	26	1874	Etta M. West
Oct.	5	1874	Austin Edgerton

# Record Linkage



Same ancestors?



# Record Linkage (continued)

- The professional genealogist finds all records about an ancestor
  - How can we do the same, linking records programmatically?
  - Same problem faced by family historians, law enforcement, credit agencies, the healthcare industry, and the CIA
- Best practices
  - Family handprint
  - Data normalization
  - Historical demographics

# Data Normalization

- Fundamental to good record linkage
- Proximity measures needed for comparing names, dates, places and events
  - Most algorithms fall short
  - Must deal with phonetically similar spellings, keyboard errors, misread letters, translated names, abbreviations, local naming patterns, and much more

# Name Normalization

- Name authorities for given names and surnames
- Authorities should be sensitive to locality, ethnicity and religion
- Most algorithms fall short
- Exhaustive solutions exists for some locales
- Need to build knowledge base for more regions
- Can we produce algorithms to improve or replace table-driven name authorities?

# Place Normalization

- Place authorities should handle variant spellings and abbreviations
- Multiple jurisdictional hierarchies
- Coordinates and boundaries
- Neighbors and nesting
- Changes over time
- What is the best way to measure proximity?

# Event Normalization

- Event hierarchies
- Event affinities
- Relationship between events and record types
  - Which events are captured in a record type?
  - Quality measures for events:
    - Original versus derivative source
    - Primary versus secondary information
    - Direct versus indirect evidence
- How should event types be characterized to provide the best measure of proximity?

# Research Guidance

- We need technology solutions to guide researchers to answers for these questions:
  - What has already been done?
  - Who should I work on now?
  - What information do I want to learn?
  - How do I get this information?
  - How do I evaluate the information?
  - What evidence did I find?
  - Do I have an answer to my question?
  - How can I be sure?

# Collaboration

- Current best practices on the Internet
  - Message boards
  - Record annotations
  - Submitter contact information
  - Personal websites
- Possible improvements
  - Shared research area for relatives
  - Tools to compare research, records and conclusions
  - Tools to assign tasks and avoid duplication of work

# Cultural Differences

- Ancestors, descendants, lineage or tribe?
- Issues viewing and navigating
- Issues with research
- Name conventions
- Locality conventions
- How to best accommodate these issues in the user interface

# Research Model

- Historical records can overwhelm the novice
- Possible solutions:
  - Just-in-time guidance
  - Highlight key fields
  - Handwriting help
  - Links to sources
  - Background info
  - Language help
  - Next steps

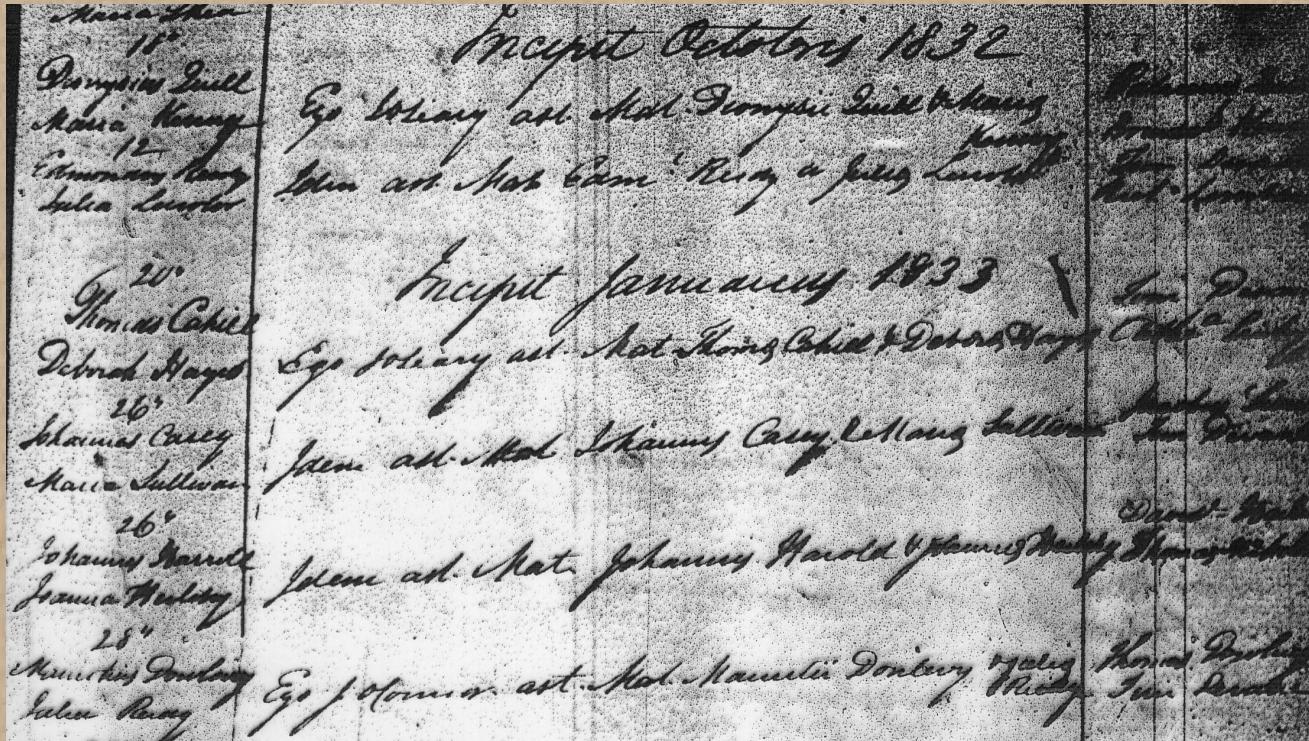
1 PLACE OF DEATH (Dist. No.) <small>(To be inserted by Registrar)</small>		West Virginia State Department of Health DIVISION OF VITAL STATISTICS	
County. <i>Magnolia</i>	District or <i>Charleston</i>	CERTIFICATE OF DEATH	
Town or City. <i>Charleston</i>	No. <i>1110 Quarrier</i>	Register No. <i>2978</i>	St.;
2 FULL NAME <i>Anna Bright Miller</i>			
PERSONAL AND STATISTICAL PARTICULARS			
3 SEX <i>F</i>	4 COLOR OR RACE <i>White</i>	5 SINGLE, MARRIED, WIDOWED, OR DIVORCED, <small>(Write the word)</small>	<i>Married</i>
6 DATE OF BIRTH <i>Dec. 30, 1856</i>		IF LESS than 1 day, _____ hrs. or _____ min?	
7 AGE <i>63 yrs / mon. 22 da.</i>	MEDICAL CERTIFICATE OF DEATH		
10 DATE OF DEATH <i>Feb. 22 - 1920</i>			11 DATE OF DEATH <i>Feb. 21, 1920 to Feb. 22 - 1920</i>
12 I HEREBY CERTIFY, That I attended deceased from <i>Feb. 21, 1920 to Feb. 22 - 1920</i> , that I last saw her alive on <i>Feb. 22 - 1920</i> , and that death occurred, on the date stated above, at <i>7:20 p.m.</i> The CAUSE OF DEATH was as follows; <i>Aphexy 64</i>			
13 CONTRIBUTORY <small>(Secondary)</small>			
14 NAME OF FATHER <i>J.C. Bright</i>			
15 BIRTHPLACE OF FATHER <small>(State or country)</small> <i>Ohio</i>			
16 MAIDEN NAME OF MOTHER <i>Ann M. Stover</i>			
17 BIRTHPLACE OF MOTHER <small>(State or country)</small> <i>Ohio</i>			
18 LENGTH OF RESIDENCE (FOR HOSPITALS, INSTITUTIONS, TRANSIENTS, OR RECENT RESIDENTS)			
At place of death <i>W. N. Miller</i> <small>(Informant)</small>		In the <i>Charleston W. Va.</i> <small>(Address)</small>	Where was disease contracted, <small>If not at place of death?</small>
Former or usual residence...			
19 PLACE OF BURIAL OR REMOVAL <i>Parkersburg W. Va.</i> <small>(Address)</small>		DATE OF BURIAL <i>Feb. 24 - 1920</i>	
20 UNDERTAKER <i>C.A. Simpson</i> <small>(Name)</small>			
ADDRESS <i>City</i>			

# Digital Image Conversion

- To convert all 2.3 million rolls of microfilm at the Granite Mountain Records Vault to digital images:
  - At the rate of one roll per hour, 8 hours a day, 5 days a week, would take over 1000 years
  - At only 1MB per image, would take 2-3 petabytes of storage space
- Need to:
  - Increase scanning capacity dramatically
  - Prioritize digital conversion
  - Improve image compression technology

# Digital Image Enhancement

- Suppress noise and bleed through
- Improve contrast, legibility
- Require little to no monitoring



# Better Transcriptions

- Current best practices:
  - Double-blind keying with arbitration
  - Statistical audits
- Manually intensive, expensive
- Technology challenges:
  - Interpret document layout
  - Recognize handwriting
  - Guide transcribers
  - Supplement or replace manual transcription

Número de orden	A CUAL ES SU APELLIDO? NOMBRE?		B	C	D	E
			Es varón o mujer	Cuantos años ha cumplido	Es soltero, casado o viudo	A qué nación pertenece
1	Sarati Juana		M.	19	S. Argentina	
2	Escobar Fidela		M.	40	S. Argentina	
3	Escobar Silvia		M.	16	S. Argentina	
4	Escobar Susetina		M.	9	S. Argentina	
5	Escobar Paulino		M.	6	S. Argentina	
6	Escobar Transito		M.	2	S. Argentina	
7	Escobar Florentina		M.	2	S. Argentina	
8	Balenuela Esulalia		M.	46	S. Argentina	
9	Balenuela Esturija		M.	24	S. Argentina	
10	Balenuela Gregorio		M.	19	S. Argentina	
11	Balenuela Foley		M.	16	S. Argentina	
12	Balenuela Martin		M.	12	S. Argentina	
13	Balenuela Maria		M.	6	S. Argentina	
14	Bego Leonzo		M.	8	S. Argentina	
15	Familan Antonia		M.	35	S. Paraguay	

# Handwriting Recognition

- Detect letters
- Detect words
- Compare the curvature of similar words
- Assist in validating transcribed text

The image shows a list of names and their relationships, handwritten in cursive script on white paper with blue horizontal lines. The list includes:

- Quimette, Charlotte Head.
- , Eleanor Wife H
- , Richard D Son
- 3 Gaudette, Ely J. Head
- , Mary K. Wife H
- , Richard S Son
- 4 Tapiras, William Head
- , Mary Wife H
- , Gladys daughter
- , Evelyn daughter
- Breslow, Julia sister law
- Rugalski, William Father-law
- Beezony, Anthony Brother-law

# Better OCR Technologies

- Detect field locations in historical forms
- Interpret values using field-sensitive vocabularies

**UNITED STATES OF AMERICA**

OATH OF ALLEGIANCE

**PETITION FOR NATURALIZATION**

Sworn before me this \_\_\_\_\_ day of \_\_\_\_\_, anno Domini \_\_\_\_\_, and I do solemnly declare and affirm, under the pains and penalties of perjury, that the following are true, to the best of my knowledge and belief.

I, William O'Connor, hereby file, respectfully commence,

First. My place of residence is 41 Ionia Street, Springfield, Massachusetts (Give number, street, city or town, and State.)

Second. My occupation is salesman.

Third. I was born on the 10th day of August, anno Domini 1888, at Tralee, Co. Kerry, Ireland.

Fourth. I immigrated to the United States from Queenstown, New York, N.Y., on or about the 1st day of July, anno Domini 1913, and arrived in the United States, at the port of New York, N.Y., on the 13th day of July, anno Domini 1913, on the vessel unknown.

(If the alien arrived otherwise than by vessel, the character of conveyance or name of transportation company should be given.)

Fifth. I declared my intention to become a citizen of the United States on the NONE day of             , anno Domini             .  
I further declare that I made no affidavit to said clerk concerning said intention, and that I have no record of having made any affidavit to any other officer or body, or signed any declaration in writing, concerning said intention.

Sixth. I am not married. My wife's name is             ; she was born on the              day of             , anno Domini             , and now resides at             . (Give number, street, city or town, and State.)

Seventh. The name, date, and place of birth, and place of residence of each of said children is as follows:

Petr. is an Honorable Discharged Soldier and files this petition under the provisions of the Act of June 29, 1906 as amended by the Act of May 2, 1918 and July 19, 1919.

Deau Park Dr. No. 1 West 9, 1919.

# Metadata Taxonomy

- Most genealogy records are not indexed by name
- Need to offer fielded access to unindexed records by:
  - Place
  - Date range
  - Event type
  - Record type

Title	Registres paroissiaux, 1797-1877
Authors	<a href="#">Eglise catholique. Saint-Jean-Baptiste-de-Rouville (Saint-Jean-Baptiste, Québec)</a> (Main Author)
Notes	Microfilm des originaux dans le Presbytère de St.-Jean-Baptiste. Comprend des baptêmes, mariages et sépultures. Index au front. Baptisms, marriages, burials.
Subjects	<a href="#">Quebec, Rouville, Saint-Jean-Baptiste - Church records</a>

Format	Manuscript (On Film)
Language	French
Publication	Salt Lake City : Filmés par la Genealogical Society of Utah, 1980
Physical	3 bobines de microfilm ; 35 mm

Title	Registres paroissiaux, 1797-1877
Authors	<a href="#">Eglise catholique. Saint-Jean-Baptiste-de-Rouville (Saint-Jean-Baptiste, Québec)</a> (Main Author)

Note	Location Film
Index, 1797-1877; registres, 1797-1819	FHL US/CAN Film 1294718
Registres, 1819-1840	FHL US/CAN Film 1294719
Registres, 1840-1877	FHL US/CAN Film 1294720

# Metadata Mining

- Goal
  - Offer access to all the world's records
- Challenge
  - Inconsistent metadata elements and schemas
- Technology problems
  - Finding genealogical collections across the Internet
  - Interpreting the metadata properly
  - Ranking the collections to best serve researchers

# Family Reconstruction

- Transcribe birth and marriage records
- Identify potential siblings
  - Manual identification
  - Automated family reconstruction

Baptismal Register's District		Births Registered in the District of		in the U		Registrar's D	
		County of Dublin					
No.	Date and Place of Birth.	Name (Mony).	Sex.	Name and Surname and Dwelling Place of Father.	Name and Surname and Maiden Surname of Mother.	Rank or Profession of Father.	Signature, C Residence
175	25th Jan 1885 St. Stephens Green	Maurice	male	Maurice Reilly	Anastasia Reilly	Coast Guard	In K Rha Ireland 380m
176	2nd Jan 1885 Corkville Parliament	William	male	John Butler	Lisette Mary Yeats	Law Student	J.W. da 1885

# Conclusion

- Family history has many challenging technology problems
- Solutions will benefit multiple industries
- Solutions will help many more people find their ancestors
- The LDS Family and Church History Department is actively addressing these technology challenges
- We welcome partners in this great work