

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

David Ouimette

25 March 2004

# Which Ancestors Can I Find Online?

Pierre Ouimette	Poussant Ouimet		
David Ouimet	Celeste Guerin		
Luce Laurin	Joseph Laurin		
Charles Ouimette	Agathe Desnoyers		
	Pierre Robert Lafontaine	Joseph Robert	
	Cordelia Robert Lafontaine	Euphrosine Frechet	Euphrosine Huot
Richard Ouimette	Marie Sausse Comtois		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

Records for These  
Ancestors Are  
Online

Pierre Ouimette	Toussaint Ouimet		
David Ouimet	Celeste Guerin		
Euse Laurin	Joseph Laurin		
Charles Ouimette	Agathe Desnoyers		
Pierre Robert Lafontaine	Joseph Robert		
Cordelia Robert Lafontaine	Euphrosine Frechet	Euphrosine Huot	Alexis Huot
Richard Ouimette	Marie Sausse Comtois		
	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 Guimet
David Ouimet	Celeste Guerin
Luce Laurin	Joseph Laurin
Charles Ouimette	Agathe Desnoyers
Pierre Robert Lafontaine	Joseph Robert
Cordelia Robert Lafontaine	Euphrosine Frechet
Richard Ouimette	Euphrosine Huot
Marie Sausse Comtois	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

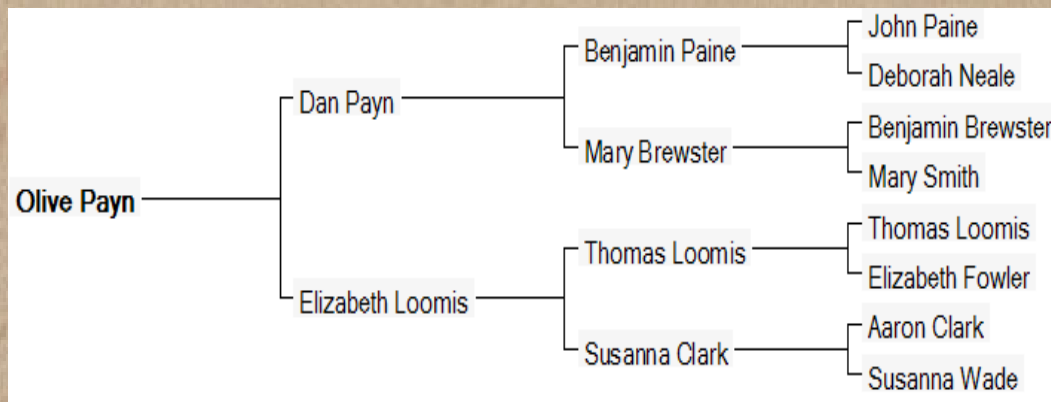


# Family History Research Problems with Records and Research

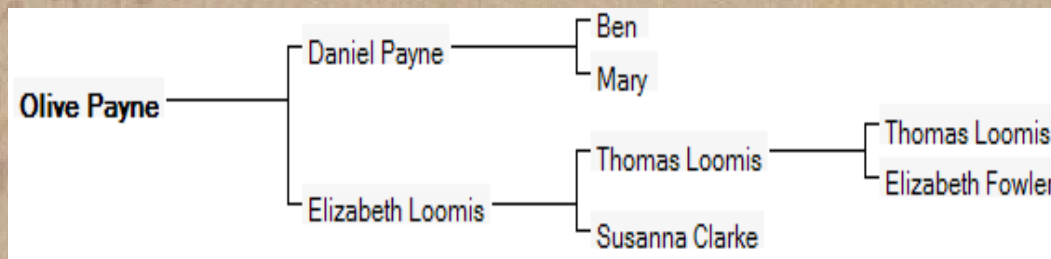
- Research model
- Digital image conversion
- Digital image enhancement
- Better transcriptions
- Handwriting recognition
- Better OCR technologies
- Metadata taxonomy
- Metadata mining
- Family reconstitution

DATE OF DEATH.			FULL NAME OF THE DECEASED.
MONTH.	DAY.	YEAR.	
June	1	1874	Lopitolia Auctin ✓
May	9	1874	John B. Prun ✓
Apr.	10	1874	Blarane Prun ✓
Oct.	20	1874	Francis Pasol ✓
Jan.	25	1874	Theolacia Whitlaw ✓
Apr.	9	1874	Alaud. Lane ✓
Dec.	18	1874	Hariland. Thayer ✓
Apr.	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. _____) (To be inserted by Registrar) County <u>Monroe</u> District <u>Charleston</u> or Town or City <u>Charleston</u>		West Virginia State Department of Health DIVISION OF VITAL STATISTICS CERTIFICATE OF DEATH Register No. <u>2978</u>	
2 FULL NAME <u>Anna Bright Miller</u>		No. <u>1110 Quarrier</u> St.;	
PERSONAL AND STATISTICAL PARTICULARS			
3 SEX <u>F</u>	4 COLOR OR RACE <u>White</u>	5 SINGLE, MARRIED, WIDOWED, OR DIVORCED <u>Married</u> (Write the word)	
6 DATE OF BIRTH <u>Dec. 30, 1856</u> (Month) (Day) (Year)			
7 AGE <u>63</u> yrs. <u>1</u> mos. <u>22</u> ds.		IF LESS than 1 day ____ hrs. or ____ min.?	
8 OCCUPATION (a) Trade, profession, or particular kind of work. (b) General nature of industry, business, or establishment in which employed (or employer). <u>Housewife</u>			
9 BIRTHPLACE (State or country) <u>Franklin Co. Ohio</u>			
PARENTS	10 NAME OF FATHER <u>J.C. Bright</u>		
	11 BIRTHPLACE OF FATHER (State or country) <u>Ohio</u>		
	12 MAIDEN NAME OF MOTHER <u>Ann M. Storer</u>		
13 BIRTHPLACE OF MOTHER (State or country) <u>Ohio</u>			
14 THE ABOVE IS TRUE TO THE BEST OF MY KNOWLEDGE (Informant) <u>W. N. Miller</u> (Address) <u>1110 Quarrier St. City</u>			
15 Filed <u>2-22</u> , 19 <u>20</u> <u>Chelard</u> REGISTRAR.			
MEDICAL CERTIFICATE OF DEATH			
16 DATE OF DEATH <u>Feb. 22</u> - <u>1920</u> (Month) (Day) (Year)			
I HEREBY CERTIFY, That I attended deceased from <u>Feb. 21</u> , 19 <u>20</u> to <u>Feb. 22</u> - 19 <u>20</u> , that I last saw <u>her</u> alive on <u>Feb. 22</u> - 19 <u>20</u> , and that death occurred, on the date stated above, at <u>7:20</u> a.m. The CAUSE OF DEATH was as follows: <u>Apoplexy</u> <u>64</u> (Duration) ____ yrs. ____ mos. ____ ds.			
CONTRIBUTORY (Secondary) ____ (Duration) ____ yrs. ____ mos. ____ ds.			
(Signed) <u>R. A. DeLund</u> , M. D. <u>2-22</u> - <u>1920</u> (Address) <u>Charleston W. Va.</u>			
NOTE: State the DISEASE CAUSING DEATH. In deaths from Violent Causes, State Means, or Injury; and whether Accidental, Suicidal, or Homicidal.			
18 LENGTH OF RESIDENCE (FOR HOSPITALS, INSTITUTIONS, TRANSIENTS, OR RECENT RESIDENTS) At place of death ____ yrs. ____ mos. ____ ds. In the ____ yrs. ____ mos. ____ ds. State ____			
Where was dissection contracted, if not at place of death?			
Former or usual residence.			
19 PLACE OF BURIAL OR REMOVAL <u>Parkersburg W. Va.</u>		DATE OF BURIAL <u>Feb. 24</u> - <u>1920</u>	
20 UNDERTAKER <u>R. A. Simpson</u>		City <u>City</u>	
ADDRESS.			

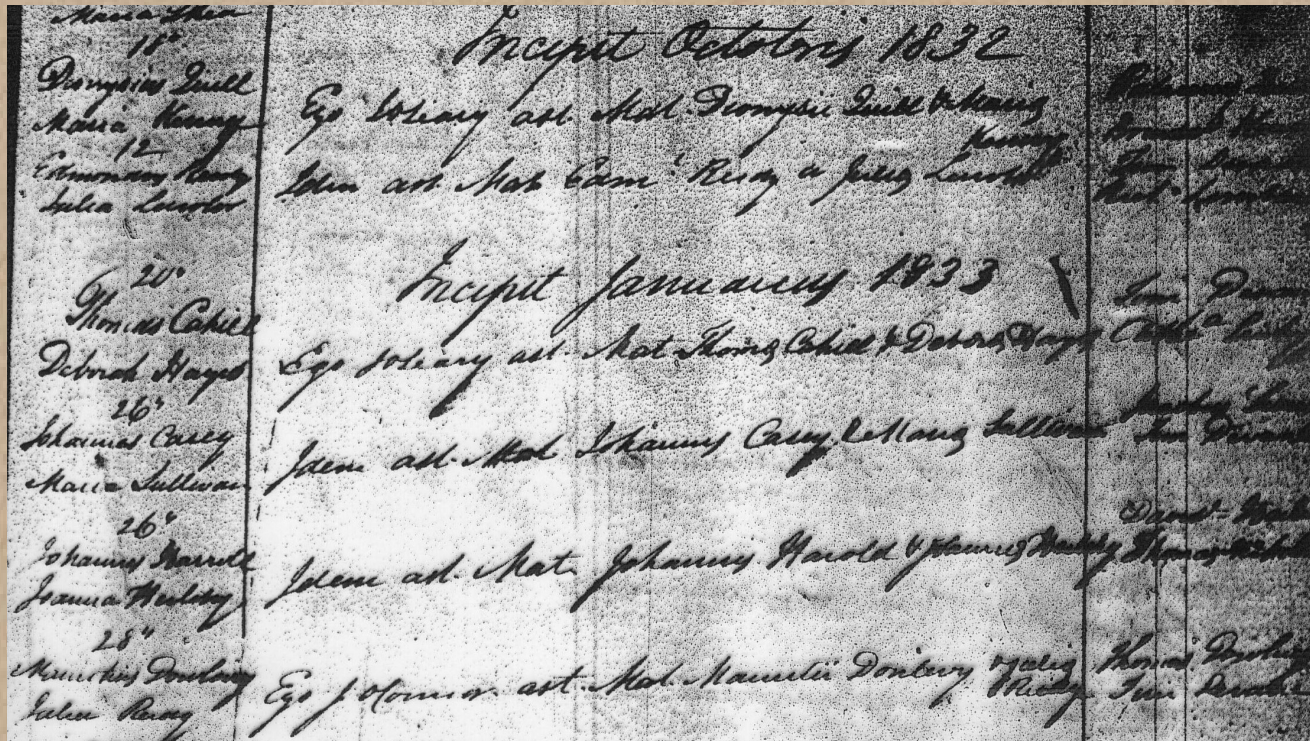
# Digital Image Conversion

- 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		B	C	D	E
	CUAL ES SU		Es varón ó mujer	Cuantos años ha cumplido	Es soltero, casado ó viudo	A qué nación pertenece
	APELLIDO?	NOMBRE?				
1	Larot	Juana	M.	19	S.	Argentina
2	Escobar	Fidela	M.	40	S.	Argentina
3	Escobar	Silvia	M.	16	S.	Argentina
4	Escobar	Asustina	M.	9		Argentina
5	Escobar	Pauline	M.	6		Argentina
6	Escobar	Tranquilo	M.	2		Argentina
7	Escobar	Florentina	M.	2		Argentina
8	Balensuela	Estelita	M.	40	S.	Argentina
9	Balensuela	Eduvigi	M.	24	S.	Argentina
10	Balensuela	Eugenio	M.	19	S.	Argentina
11	Balensuela	Eloy	M.	16	S.	Argentina
12	Balensuela	Martin	M.	12		Argentina
13	Balensuela	Maria	M.	6		Argentina
14	Bega	Louisa	M.	5		Argentina
15	Garcilán	S. Antonio	M.	35	S.	Paraguay



# Handwriting Recognition

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





# Better OCR Technologies

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

*Filed  
Superior  
Court  
8/12*

**UNITED STATES OF AMERICA**

**OATH OF ALLEGIANCE**

**PETITION FOR NATURALIZATION**

The petition of William O'Connor of the County of Hampden State of Massachusetts hereby filed, respectfully sheweth:

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 1 888, at Tralee, Co. Kerry, Ireland

Fourth. I emigrated to the United States from Queenstown, New York, N. Y., on or about the 1st day of July

Domini 1 913 and arrived in the United States, at the port of New York, N. Y., on the 13th day of July

Domini 1 913, 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 \_\_\_\_\_ day of \_\_\_\_\_, anno Domini 1 \_\_\_\_\_

in the \_\_\_\_\_ Court of \_\_\_\_\_

Sixth. I am not married. My wife's name is \_\_\_\_\_; she was born on the \_\_\_\_\_ day of \_\_\_\_\_, anno Domini 1 \_\_\_\_\_

\_\_\_\_\_ and now resides at \_\_\_\_\_  
(Give number, street, city or town, and State.)

children, and the name, date, and place of birth, and place of residence of each of said children is as follows:

Petr. is an Honorably Discharged Soldier and files this petition under the  
visions of the Act of June 29, 1906 as amended by the Act of May 9, 1918 and July 19, 1919  
See Case No. 12-1919

# 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	<p>Microfilm des originaux dans le Presbytère de St.-Jean-Baptiste.</p> <p>Comprend des baptêmes, mariages et sépultures.</p> <p>Index au front.</p> <p>Baptisms, marriages, burials.</p>
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

Superintendent Registrar's District <i>North Dublin Union</i>				Registrar's District			
Births Registered in the District of <i>Donnybrook</i>				in the Union of <i>County of Dublin</i>			
No.	Date and Place of Birth.	Name (if any).	Sex.	Name and Surname and Dwelling Place of Father.	Name and Surname and Maiden Surname of Mother.	Rank or Profession of Father.	Signature, Residence.
195	<i>21st Dec 1880</i> <i>Black Horse</i> <i>Donnybrook</i>	<i>Maurice</i>	<i>male</i>	<i>Maurice</i> <i>Reilly</i> <i>Ringsend</i>	<i>Anastasia</i> <i>Reilly</i> <i>James</i> <i>McCarthy</i>	<i>Constable</i>	<i>M. K.</i> <i>Reilly</i> <i>1890</i>
196	<i>21st Dec 1880</i> <i>George Hill</i> <i>Donnybrook</i>	<i>William</i> <i>Butler</i>	<i>male</i>	<i>John</i> <i>Butler</i> <i>Yeats</i> <i>George Hill</i>	<i>Susan</i> <i>Mary</i> <i><del>Butler</del></i> <i>Yeats</i> <i>James</i> <i>Pollard</i>	<i>Law Student</i>	<i>J. B.</i> <i>Butler</i> <i>1890</i>

# 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