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

David Ouimette
25 March 2004
<table>
<thead>
<tr>
<th>Pierre Ouimette</th>
<th>Toussaint Ouimet</th>
</tr>
</thead>
<tbody>
<tr>
<td>David Ouimet</td>
<td>Celeste Guerin</td>
</tr>
<tr>
<td>Luc Laurin</td>
<td>Joseph Laurin</td>
</tr>
<tr>
<td>Charles Ouimet</td>
<td>Agathe Desnoyers</td>
</tr>
<tr>
<td>Pierre Robert Lafontaine</td>
<td>Joseph Robert</td>
</tr>
<tr>
<td>Cordelia Robert Lafontaine</td>
<td>Euphrosine Frechet</td>
</tr>
<tr>
<td></td>
<td>Euphrosine Huot</td>
</tr>
<tr>
<td>Richard Ouimet</td>
<td>Marie Sausse Comtois</td>
</tr>
<tr>
<td></td>
<td>Jozef Czelusniak</td>
</tr>
<tr>
<td></td>
<td>Grzegorz Czelusniak</td>
</tr>
<tr>
<td></td>
<td>Stanislaw Czelusniak</td>
</tr>
<tr>
<td></td>
<td>Katarzyna Krupa</td>
</tr>
<tr>
<td></td>
<td>Katarzyna Trzaskos</td>
</tr>
<tr>
<td></td>
<td>Walenty Trzaskos</td>
</tr>
<tr>
<td></td>
<td>Eleanora Czelusniak</td>
</tr>
<tr>
<td></td>
<td>Agata Dybas</td>
</tr>
<tr>
<td></td>
<td>Jakub Jarek</td>
</tr>
<tr>
<td></td>
<td>Jan Jarek</td>
</tr>
<tr>
<td></td>
<td>Katarzyna Jarek</td>
</tr>
<tr>
<td></td>
<td>Katarzyna Bujak</td>
</tr>
<tr>
<td></td>
<td>Anna Czelusniak</td>
</tr>
<tr>
<td></td>
<td>Jan Czelusniak</td>
</tr>
<tr>
<td></td>
<td>David Scott Ouimette</td>
</tr>
<tr>
<td></td>
<td>Marianna Dybas</td>
</tr>
<tr>
<td></td>
<td>John Connor</td>
</tr>
<tr>
<td></td>
<td>Michael O’Connor</td>
</tr>
<tr>
<td></td>
<td>William O’Connor</td>
</tr>
<tr>
<td></td>
<td>Ambrose Lysaght Browne</td>
</tr>
<tr>
<td></td>
<td>Bartholomew Browne</td>
</tr>
<tr>
<td></td>
<td>Mary Jane Browne</td>
</tr>
<tr>
<td></td>
<td>Margaret Lysaght</td>
</tr>
<tr>
<td></td>
<td>Mary West</td>
</tr>
<tr>
<td></td>
<td>Thomas West</td>
</tr>
<tr>
<td>Marilyn O’Connor</td>
<td></td>
</tr>
<tr>
<td></td>
<td>Marble Bissonette</td>
</tr>
<tr>
<td></td>
<td>Amable Bissonnet</td>
</tr>
<tr>
<td></td>
<td>Alfred Marble Bissonette</td>
</tr>
<tr>
<td></td>
<td>Phebe Battiste</td>
</tr>
<tr>
<td></td>
<td>Clarissa White</td>
</tr>
<tr>
<td></td>
<td>Newell White</td>
</tr>
<tr>
<td>Merle Carlyss Bissonett</td>
<td>Clarissa Mills</td>
</tr>
<tr>
<td></td>
<td>Edward Tallman</td>
</tr>
<tr>
<td></td>
<td>Joseph Tallman</td>
</tr>
<tr>
<td>Blanche Elvina Tallman</td>
<td>Jane Patterson</td>
</tr>
<tr>
<td>Clarissa Wade</td>
<td>Horatio Wade</td>
</tr>
<tr>
<td>Phebe Tarte</td>
<td></td>
</tr>
</tbody>
</table>
Records for These Ancestors Are Online
These Ancestors Are Found Online or at the FHL
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
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
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
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
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

---

<table>
<thead>
<tr>
<th>Title</th>
<th>Registres paroissiaux, 1797-1877</th>
</tr>
</thead>
<tbody>
<tr>
<td>Authors</td>
<td>Église catholique, Saint-Jean-Baptiste-de-Rouville (Saint-Jean-Baptiste, Québec) (Main Author)</td>
</tr>
<tr>
<td>Subjects</td>
<td>Quebec, Rouville, Saint-Jean-Baptiste - Church records</td>
</tr>
</tbody>
</table>

---

<table>
<thead>
<tr>
<th>Format</th>
<th>Manuscript (On Film)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Language</td>
<td>French</td>
</tr>
<tr>
<td>Publication</td>
<td>Salt Lake City : Filmés par la Genealogical Society of Utah, 1980</td>
</tr>
<tr>
<td>Physical</td>
<td>3 bobines de microfilm ; 35 mm</td>
</tr>
</tbody>
</table>

---

<table>
<thead>
<tr>
<th>Note</th>
<th>Location</th>
</tr>
</thead>
<tbody>
<tr>
<td>Index, 1797-1877; registres, 1797-1819</td>
<td>FHL US/CAN Film 1294718</td>
</tr>
<tr>
<td>Registres, 1819-1840</td>
<td>FHL US/CAN Film 1294719</td>
</tr>
<tr>
<td>Registres, 1840-1877</td>
<td>FHL US/CAN Film 1294720</td>
</tr>
</tbody>
</table>
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
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