Research in the Acquisition Pipeline

David W. Embley

Christopher Almquist, Bill Barrett, Alan Cannaday, Robert Clawson, Jake Gehring, Doug Kennard, Tae Woo Kim, Steve Liddle, Peter Lindes, Deryle Lonsdale, Thomas Packer, Joseph Park, Pat Schone, Scott Woodfield
Acquisition Pipeline

Strategic Planning
Field Negotiations
\underline{Field Capture}
HQ Image & Metadata Ingest
Image Auditing
Cataloging
Collection Treatment
Waypointing
Book Scanning
Oral History Recording
\underline{Indexing}
Post-Processing/Quality Control
\underline{Load to Search Engine, Publish}
Field Capture: Blur Detection
(Alan Cannaday)
Field Capture: Blur Detection

Sharp/Focused  Out of Focus  Motion Blur
Field Capture: Blur Detection

- Sharp/Focused
- Out of Focus
- Motion Blur

“More than two transitional pixels between the edge of a high contrast line and the background.”

“Transitional pixels in a single direction exceed one pixel.”
Field Capture: Blur Detection

- Failed: 82.0%
- Passed: 83.5%

![Graph showing normalized amplitude vs. mean growth rate with labeled categories like Pass, Fail, Blur, Out of Focus, and their smoothed versions.](Image)
Load to Search Engine: Constraint Satisfaction (Scott Woodfield)
Existing assertions:
Blood type of father, mother, and child all A-.
New assertion:
Child’s blood type B-

Conclusions: Probability = 0.0.
(1) Parentage wrong
(2) One or more blood types wrong
Automated “Green” Indexing

“Green”: improves with use—learns from user interaction

- Intelligent Indexing
- “Click” Annotator
- GreenFIE-HD
- Obituaries (100M+)
  - FROntIERER
  - Machine Learning
- Scanned Books (100K+)
  - ListReader
  - FormReader/TableReader
  - OntoSoar
- GreenFIE-HD++
“Green” Intelligent Indexing
(Robert Clawson, Doug Kennard, ..., Bill Barrett)
“Green” Intelligent Indexing
“Green” Intelligent Indexing
“Green” Intelligent Indexing
THE ELY ANCESTRY.

SEVENTH GENERATION.

241213. Mary Eliza Warner, b. 1826, dau. of Samuel Selden Warner and Azubah Tully; m. 1850, Joel M. Gloyd (who was connected with Chief Justice Waite's family).

243311. Abigail Huntington Lathrop (widow), Boonton, N. J., b. 1810, dau. of Mary Ely and Gerard Lathrop; m. 1835, Donald McKenzie, West Indies, who was b. 1812, d. 1839.

(The widow is unable to give the names of her husband's parents.)

Their children:

1. Mary Ely, b. 1836, d. 1839.
2. Gerard Lathrop, b. 1838.

243312. William Gerard Lathrop, Boonton, N. J., b. 1812, d. 1882, son of Mary Ely and Gerard Lathrop; m. 1835, Charlotte Brackett Jennings, New York City, who was b. 1818, dau. of Nathan Tilestone Jennings and Maria Miller. Their children:

1. Maria Jennings, b. 1838, d. 1840.
2. William Gerard, b. 1840.
3. Donald McKenzie, b. 1840, d. 1843. [Twins].
4. Anna Margareta, b. 1845.
5. Anna Catherine, b. 1845.

243314. Charles Christopher Lathrop, N. Y. City, b. 1817, d. 1865, son of Mary Ely and Gerard Lathrop; m. 1836, Mary Augusta Andruss, 992 Broad St., Newark, N. J., who was b. 1825, dau. of Judge Caleb Halstead Andruss and Emma Sutherland Golble. Mrs. Lathrop died at her home, 992 Broad St., Newark, N. J., Friday morning, Nov. 4, 1898. The funeral services were held at her residence on Monday, Nov. 7, 1898, at half-past two o'clock P. M. Their children:

1. Charles Halstead, b. 1837, d. 1861.
2. William Gerard, b. 1838, d. 1860.
THE ELY ANCESTRY.

419

419

243311. Abigail Huntington Lathrop (widow), Boonton, N. J., b. 1810, dau. of Mary Ely and Gerard Lathrop; m. 1835, Donald McKenzie, West Indies, who was b. 1812, d. 1839.

(The widow is unable to give the names of her husband's parents.)

Their children:

1. Mary Ely, b. 1835, d. 1839.
2. Gerard Lathrop, b. 1838.

243312. William Gerard Lathrop, Boonton, N. J., b. 1812, d. 1882, son of Mary Ely and Gerard Lathrop; m. 1837, Charlotte Brackett Jennings, New York City, who was b. 1818, dau. of Nathan Tilestone Jennings and Maria Miller. Their children:

1. Maria Jennings, b. 1838, d. 1840.
2. William Gerard, b. 1840.
3. Donald McKenzie, b. 1840, d. 1843. {Twins.
4. Anna Marguerite, b. 1843.
5. Anna Catherine, b. 1845.

243314. Charles Christopher Lathrop, N. Y. City, b. 1817, d. 1865, son of Mary Ely and Gerard Lathrop; m. 1856, Mary Augusta Andrews, 952 Broad St., Newark, N. J., who was b. 1835, dau. of Judge Caleb Halstead Andrews and Emma Sutherland Gobie. Mrs. Lathrop died at her home, 992 Broad St., Newark, N. J., Friday morning, Nov. 4, 1898. The funeral services were held at her residence on Monday, Nov. 7, 1898, at half-past two o'clock P. M. Their children:

2. William Gerard, b. 1858, d. 1860.
GreenFIE-HD (Tae Woo Kim)

“Green” Form-based Information Extraction for Historical Documents
GreenFIE-HD: Extraction Rule Creation

1. Maria Jennings, b. 1838, d. 1840.
2. William Gerard, b. 1840.
3. Donald McKenzie, b. 1840, d. 1843. { Twins.
4. Anna Margareta, b. 1843.
5. Anna Catherine, b. 1845.

\d{1}\s\([A-Z][a-z]{2,6}\)\s\([A-Z][a-z]{4,10}\),\sb\s\d{4},\sd\s\d{4}\.}
GreenFIE-HD: Recall Error Resolution

2. William Gerard, b. 1858, d. 1861.
3. Theodore Andruss, b. 1860
4. Emma Goble, b. 1862.

\d{1}\s([A-Z][a-z]{2,8})\s([A-Z][a-z]{1,8})\sb\s(d{4})(\.|\sd\s(d{4}))

\d{1}\s([A-Z][a-z]{2,8})\s([A-Z][a-z]{1,8})\sb\s(d{4}|i\d{3})(\.|\sd\s(d{4}))
GreenFIE-HD: Precision Error Resolution

members from the old colonial Aldens, Brewsters, Fullers, and Elys. Home Journal, Jan. 15, 1896. 243316. Mary Cornelia Lathrop, b. 1820, d. 1879, dau. of Mary Ely and Gerard Lathrop; m. 1840, Jacob Dykeman Vermilye, 49 East 21st St., New York City, who was b. 1817, son of William Vermilye and Mary Montgomery. Their children
1. William Gerard, b. 1841.
2. Mary Ann, b. 1846.
3. Louisa Lathrop, b. 1851, d. 1853.

\s([A-Z][a-z]{2,8})\s([A-Z][a-z]{1,8}),
\d{1}\s ([A-Z][a-z]{2,8})\s([A-Z][a-z]{1,8}),\sb\s
GreenFIE-HD: Principles

• Look-ahead: automatic extraction

• Look-behind: rule derivation and adjustment

• “Green”: improves with use
Obituaries with FRONtIER (Joseph Park)

(Fact Recognizer for Ontologies with Inference and Entity Resolution)

Brian Fielding Frost
Our beloved Brian Fielding Frost, age 41, passed away Saturday morning, March 7, 1998, due to injuries sustained in an automobile accident. He was born August 4, 1956 in Salt Lake City, to Donald Fielding and Helen Glade Frost. He married Susan Fox on June 1, 1981. He is survived by Susan; sons Jordan (9), Travis (8), Bryce (6); parents, three brothers, Donald Glade (Lynne), Kenneth Wesley (Ellen), Alex Reed, and two sisters, Anne (Dale) Elkins and Sally (Kent) Britton. A son, Michael Brian Frost, preceded him in death.

Funeral services will be held at 12 noon Friday, March 13, 1998 in the Howard Stake Center, 350 South 1600 East. Friends may call 5-7 p.m. Thursday at Wasatch Lawn Mortuary, 3401 S. Highland Drive, and at the Stake Center from 10:45-11:45 a.m. Friday.

Interment at Wasatch Lawn Memorial Park.
Obituaries with FROntIER

Brian Fielding Frost
Our beloved Brian Fielding Frost, age 41, passed away Saturday, March 7, 1998, due to injuries in an automobile accident. He was born August 4, 1956 in Salt Lake City, Utah, to Donald Fielding and Helen Glade Frost. He married Susan Fox on June 14, 1978. He is survived by Susan; sons, Alexander (9), Travis (8), Bryce (4); three brothers, Donald Glade Frost, Kenneth Wesley (Ellen), Alex Fielding; and two sisters, Anne (Dale) and Sally (Kent) Britton. A son, Brian Frost, preceded him in death.

Funeral services will be held at 12 noon Friday, March 13, 1998 in the Howard Stake Center, 350 South 1600 East. Friends may call 5-7 p.m. Thursday at Wasatch Lawn Mortuary, 3401 S. Highland Drive, and at the Stake Center from 10:45-11:45 a.m. Friday. Interment at Wasatch Lawn Memorial Park.
Obituaries with FROntIERER
Obituaries with FRONtIERER

//Ancestor
[(?a rdf:type s:Ancestor) -> (?a rdf:type t:Person)]
//Name
[(?n rdf:type s:AncestorName) -> (?n rdf:type t:Name)]
[(?a s:Ancestor-AncestorName ?n), (?n s:AncestorNameValue ?nv)]
->
(?a t:Person-Name ?n), (?n t:NameValue ?nv)
//Person-Child relationship for Ancestor (only for parent in this example)
[(?p rdf:type s:DeceasedPerson), (?a rdf:type s:Ancestor),
 (?a s:Ancestor-AncestorRelationship ?ar),
 (?ar s:AncestorRelationshipValue 'parent')
->
(?p rdf:type t:Child), (?a rdf:type t:Person), (?a t:Person-Child ?p)]
//Gender
[(?father rdf:type s:Ancestor), makeSkolem(?g, ?father),
 (?father s:Ancestor-AncestorName ?n),
 (?n s:AncestorNameValue ?nv),
 isMale(?nv, 0.90)
->
(?g rdf:type t:Gender), (?father t:Person-Gender ?g),
 (?g t:GenderValue 'Male')]}
Obituaries with FROntIER
Obituaries with FROntIERER

Donald Fielding Frost & Helen Glade Frost
Donald Glade & Lynn Frost
Dale & Anne Frost Elkins
Jordan Frost
Travis Frost
Michael Brian Frost
Bryce Frost
Brian Fielding & Susan Fox Frost
Alex Reed Frost
Kent & Sally Frost Britton
Kenneth Wesley & Ellen Frost
Donald Fielding Frost & Helen Glade Frost
Obituaries with Machine Learning
(Pat Schone)

Sami Stephens Pearson
May 13, 1924 ~ August 24, 2018

Sami Stephens Pearson Our loving wife and mother went home to be with Jesus after a long battle with kidney failure.

She left behind her husband, Jerry; two sons, Robby and Vince; and five grandchildren. She will be missed by all her friends and family.

A Memorial Service will be held Friday, August 24, 2018, at 2 pm at North Park Foursquare Church, 4275 S 1900, Roy, Utah.

Steven Thornton
October 14, 2018

Funeral services for Steven will be held Tuesday, November 12, 2018, 11:00 a.m., at the Orangeville LDS Ward Chapel (45 West 100 South) in Orangeville.

Family will be at Mitchell Funeral Home (233 East Main Street) in Price Monday evening from 6:00-8:00 p.m. and at the church Tuesday from 9:30-10:30 a.m. prior to services.

Interment will be in the Orangeville City Cemetery.
Obituaries with GreenFIE-HD

<table>
<thead>
<tr>
<th>DeceasedPerson</th>
</tr>
</thead>
<tbody>
<tr>
<td>Name: Brian Fielding Frost</td>
</tr>
<tr>
<td>Age: 47</td>
</tr>
<tr>
<td>DeathDate: March 7, 1998</td>
</tr>
<tr>
<td>DeathPlace:</td>
</tr>
<tr>
<td>BirthDate: August 4, 1956</td>
</tr>
<tr>
<td>BirthPlace: Salt Lake City</td>
</tr>
</tbody>
</table>

Our beloved Brian Fielding Frost, age 47, passed away Saturday morning, March 7, 1998, due to injuries sustained in an automobile accident. He was born August 4, 1956 in Salt Lake City, to Donald Fielding and Helen Glade Frost. He married Susan Fox on June 1, 1981. He is survived by Susan; sons Jordan (9), Travis (8), Bryce (6); parents, three brothers, Donald Glade (Lynne), Kenneth Wesley (Ellen), Alex Reed, and two sisters, Anne (Dale) Elkins and Sally (Kent) Britton. A son, Michael Brian Frost, preceded him in death. Funeral services will be held at 12 noon Friday, March 13, 1998 in the Howard Stake Center, 350 South 1600 East. Friends may call 5-7 p.m. Thursday at Wasatch Lawn Mortuary, 3401 S. Highland Drive, and at the Stake Center from 10:45-11:45 a.m. Friday. Interment at Wasatch Lawn Memorial Park.
1555. Elias Mather, b. 1750, d. 1788, son of Deborah Ely and Richard Mather; m. 1771, Lucinda Lee, who was b. 1752, dau. of Abner Lee and Elizabeth Lee. Their children:

1. Andrew, b. 1772.
2. Clarissa, b. 1774.
3. Elias, b. 1776.
5. Sylvester, b. 1782.
7. Charles, b. 1787.

1556. Deborah Mather, b. 1752, d. 1826, dau. of Deborah Ely and Richard Mather; m. 1771, Ezra Lee, who was b. 1749 and d. 1821, son of Abner Lee and Elizabeth Lee. Their children:

1. Samuel Holden Parsons, b. 1772, d. 1870, m. Elizabeth Sullivan.
2. Elizabeth, b. 1774, d. 1851, m. 1801 Edward Hill.
3. Lucia, b. 1777, d. 1778.
4. Lucia Mather, b. 1779, d. 1870, m. John Marvin.
5. Polly, b. 1782.
6. Phebe, b. 1783, d. 1805.
8. Margaret Stoutenburgh, b. 1794.
1555. Elias Mather, b. 1750, d. 1788, son of Deborah Ely and Richard Mather; m. 1771, Lucinda Lee, who was b. 1752, dau. of Abner Lee and Elizabeth Lee. Their children:

1. Andrew, b. 1772.
2. Clarissa, b. 1774.
3. Elias, b. 1776.
5. Sylvester, b. 1782.
7. Charles, b. 1787.

1556. Deborah Mather, b. 1752, d. 1826, dau. of Deborah Ely and Richard Mather; m. 1771, Ezra Lee, who was b. 1749 and d. 1821, son of Abner Lee and Elizabeth Lee. Their children:

1. Samuel Holden Parsons, b. 1772, d. 1870, m. Elizabeth Sullivan.
2. Elizabeth, b. 1774, d. 1851, m. 1801 Edward Hill.
3. Lucia, b. 1777, d. 1778.
4. Lucia Mather, b. 1779, d. 1870, m. John Marvin.
5. Polly, b. 1782.
6. Phebe, b. 1783, d. 1805.
8. Margaret Stoutenburgh, b. 1794.
ListReader

- [Sp] [Dg] [Sp] [UpLo+] [Sp] [Lo] [Sp] [DgDgDgDg] [Sp]
- "n1. Andrew, b. 1772.\n- "n2. Clarissa, b. 1774.\n- "n3. Elias, b. 1776.\n- "n5. Polly, b. 1782.\n- "n5. Sylvester, b. 1782.\n- "n7. Charles, b. 1787.\n- "n8. Margaret Stoutenburgh, b. 1794.\n- [Sp] [Dg] [Sp] [UpLo+] [Sp] [Lo] [Sp] [DgDgDgDg] [Sp] [Lo] [Sp] [DgDgDgDg] [Sp]
- "n4. William Lee, b. 1779, d. 1802.\n- "n6. Nathaniel Griswold, b. 1784, d. 1785.\n- "n3. Lucia, b. 1777, d. 1778.\n- "n6. Phebe, b. 1783, d. 1805.\n- "n7. William Richard Henry, b. 1787, d. 1796.\n
ListReader

(\[n\])(\[d\]{1})(.\[n\])(\[A-Z]+[a-z]+|A-Z+a-z+A-Z+a-z+)(,)(\[n\])(\[d\]{4})(.)(\[n\])

Label Fields

??

Person


<table>
<thead>
<tr>
<th>ChildNr</th>
<th>5</th>
</tr>
</thead>
<tbody>
<tr>
<td>Name</td>
<td>PoUy</td>
</tr>
<tr>
<td>BirthDate</td>
<td>1782</td>
</tr>
</tbody>
</table>
ListReader  (Thomas Packer)
GENERAL DECLARATION
(Outward/Return)

AGRICULTURE, CUSTOMS, IMMIGRATION, AND PUBLIC HEALTH

Owner or Operator: NATIONAL AERONAUTICS AND SPACE ADMINISTRATION

Arrival Ports and Stages

GEO. FREDERICK

Flight: APOLLO II
Date: JULY 24, 1969

Departure from:

(Place and Date)

FLIGHT ROUTING

("Place" Columns always list origin, every on-route stop and destination)

<table>
<thead>
<tr>
<th>PLACE</th>
<th>TOTAL NUMBER OF CREW</th>
<th>NUMBER OF PASSENGERS ON THIS STAGE</th>
<th>CARGO</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

 Declaration of Health:

Pursuant to laws known to be suffering f or illness other than sickness or the effect of accidents, as well as those of illness discovered during the flight:

NONE

Any other condition on board which may lead to the spread of disease:

TO BE DETERMINED

Details of each uncompensated or sanitary treatment (place, date, time, method) during the flight. If no compensation has been reached during the flight give details of all recent complications:

Signed, if required: Geo. F. Collins, Ground Inspector

I declare that all statements and particulars contained in this General Declaration, and in any supplementary forms required to be presented with this General Declaration are complete, exact and true to the best of my knowledge and that all through passengers will continue to travel on the flight.
ChartReader, Table Reader, ...

Donald Fielding Frost & Helen Glade Frost

Michael Brian Frost

Brian Fielding & Susan Fox Frost

Travis Frost

Dale & Anne Frost Elkins

Bryce Frost

Michael Brian Frost

Kent & Sally

Kenneth Wesley & Ellen Frost

Bryan Fielding Frost & Helen Glade Frost

Alex Reed Frost

Travis Frost
OntoSoar (Peter Lindes, Deryle Lonsdale)

Children of James Harwood, No. 103.

229. Myra, born July 26, 1835, in Eden, Vt. She married Elijah Spencer, Dec. 25, 1851. They had five children: Arvilla, born in 1852, is not living; Mariette, born Dec. 25, 1854, married Jonathan Snyder, have a family; Leverett, born Feb. 6, 1857, married Cora Smith, Nov. 2, 1879, had two children, Perry F. and Ida I. Leverett died May 21, 1910; Rosa E., born Jan. 13, 1860, married Emmett Byers, and have children; and Harrison, born about 1862, is not living. Elijah Spencer died in the Union army in 1863, and his widow married Jonathan Squires, who was born in Ohio, July 25, 1823, by whom she had one son, J. Wilbur, born June 16, 1865, in DeKalb county, Ind., married Cora M. Thomas, Aug. 24, 1887, they reside in St. Joseph, Mich., five children. Mrs. Myra Squires died in Allen county, Ind., Feb. 13, 1874.
OntoSoar (Peter Lindes, Deryle Lonsdale)

^ Mary died.v in 1853.

Person(X1)
Name(X2,"Mary")
Person(X1) has Name(X2)
DeathDate(X3,"1853")
Person(X1) died on DeathDate(X3)

Person(...) has Name(...) 
DeathDate(...) 
Person(...) died on DeathDate(...) 

4/7/2014 BYU CS Colloquium
OntoSoar  (Peter Lindes, Deryle Lonsdale)

3. Theodore Knaruss, d. 1800.
4. Emma Goble, b. 1862.

Miss Emma Goble Lathrop, official historian of the New York Chapter of the Daughters of the American Revolution, is one of the youngest members to hold office, but one whose intelligence and capability qualify her for such distinction. Miss Lathrop is not without experience; in her present home and native city, Newark, N. J., she has filled the positions of secretary and treasurer to the Girls' Friendly Society for nine years, secretary and president of the Woman's Auxiliary of Trinity Church Parish, treasurer of the St. Catherine's Guild of St. Barnabas.
OntoSoar  (Peter Lindes, Deryle Lonsdale)

^ Emma was a official historian of the NYCDAR.

Person-Officer-Organization(y1,"official historian","NYCDAR")
GreenFIE-HD++

FROntIER

ListReader

OntoSoar

Ever learning & improving
Research Wish List

Strategic Planning
Field Negotiations
Field Capture
HQ Image & Metadata Ingest
Image Auditing
Cataloging
Collection Treatment
Waypointing
Book Scanning
Oral History Recording
Indexing
Post-Processing/Quality Control
Load to Search Engine, Publish

OCR alignment with images across fonts and typesetting layouts

Automated extraction from filled-in forms, tables and ahnentafel templates

Semantic OCR error correction
Research Wish List (Jake Gehring)

- Strategic Planning
- Field Negotiations
- Field Capture
  - HQ Image & Metadata Ingest
  - Image Auditing
  - Cataloging
  - Collection Treatment
  - Waypointing
- Book Scanning
- Oral History Recording
- Indexing
- Post-Processing/Quality Control
- Load to Search Engine, Publish

Facial recognition based on labeled faces in other photos

Social/collaborative indexing environments

Snippet indexing on mobile devices
Research Wish List (Jake Gehring)

Strategic Planning
Field Negotiations
Field Capture
   HQ Image & Metadata Ingest
   Image Auditing
   Cataloging
   Collection Treatment
   Waypointing
   Book Scanning
   Oral History Recording
   Indexing
   Post-Processing/Quality Control

Search results clustering based on kinship networks

Extraction of lineage-linked data in register-style tables

Handwriting recognition
Research Wish List (Jake Gehring)

Strategic Planning
Field Negotiations
Field Capture
HQ Image & Metadata Ingest
Image Auditing
Cataloging
Collection Treatment
Waypointing
Book Scanning
Oral History Recording
Indexing
Post-Processing/Quality Control
Load to Search Engine, Publish

Extraction of lineage-linked data from text

Newspaper scanning, zoning, article concatenation

Records hinting for historical collections
Research Wish List (Jake Gehring)

- Strategic Planning
- Field Negotiations
- Field Capture
  - HQ Image & Metadata Ingest
  - Image Auditing
  - Cataloging
  - Collection Treatment
  - Waypointing
  - Book Scanning
  - Oral History Recording
  - Indexing
  - Post-Processing/Quality Control

Image capture software to eliminate blur and focus issues

Efficient routing of work to volunteers

Automatic document classification

Load to Search Engine, Publish
Summary

Streamline the Pipeline: Research Opportunities

“turn ... the heart of the children to their fathers”